



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

SEGREGACIÓN EN EDUCACIÓN MEDIA: APROXIMACIONES A TRAVÉS DE UN MODELO DE ELECCIÓN DISCRETA.

FELIPE HERNÁN ARTEAGA OSSA

Tesis para optar al grado de
Magíster en Ciencias de la Ingeniería

Profesor Supervisor:
RICARDO PAREDES

Santiago de Chile, Agosto 2014

© MMXIII, FELIPE HERNÁN ARTEAGA OSSA



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

SEGREGACIÓN EN EDUCACIÓN MEDIA: APROXIMACIONES A TRAVÉS DE UN MODELO DE ELECCIÓN DISCRETA.

FELIPE HERNÁN ARTEAGA OSSA

Tesis presentada a la Comisión integrada por los profesores:

RICARDO PAREDES

JUAN DE DIOS ORTÚZAR

VALENTINA PAREDES

JOSÉ LUIS ALMAZÁN

Para completar las exigencias del grado de
Magíster en Ciencias de la Ingeniería

Santiago de Chile, Agosto 2014

© MMXIII, FELIPE HERNÁN ARTEAGA OSSA

AMDG

AGRADECIMIENTOS

A mi familia, por confiar en que este largo proceso de tesis ha sido reflexivo y fructífero, especialmente a mi papá, por sus útiles comentarios. A mis amigos, por acompañarme y escucharme hablar de “la tesis”.

A los compañeros de la sala de Magíster de Ingeniería Industrial, por todo el tiempo compartido. Al equipo de J-PAL LAC, por acogerme en el proceso de práctica y ayudarme con reflexiones relacionadas a la tesis.

Al profesor de la Escuela de Ingeniería Juan de Dios Ortúzar, por introducirme en pregrado a los modelos de elección discreta, y por su gran colaboración en este trabajo. Al profesor del Instituto de Economía Tomás Rau, por transmitir su interés por la econometría.

A Pedro Lizana, Javier Ossa, Juan Carlos Muñoz, Maximiliano Maroto, León Délano y Benito Baranda por sus valiosos comentarios respecto del borrador final.

Por último a mis profesores guías, Valentina Paredes y Ricardo Paredes, por confiar y darme la oportunidad de desarrollar el tema, además de sus valiosas reflexiones, paciencia, apoyo y pronta respuesta.

ÍNDICE GENERAL

AGRADECIMIENTOS	IV
ÍNDICE GENERAL	V
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	XI
RESUMEN	XIII
ABSTRACT	XIV
1. INTRODUCCIÓN	1
1.1 Sistema educacional chileno	4
2. BREVE REVISIÓN BIBLIOGRÁFICA	9
2.1 Preferencias en Elección de Colegios	9
2.1.1 El caso chileno	10
2.2 Medición de Segregación	14
2.2.1 Índices de segregación para grupos dicotómicos	14
2.2.2 Índices de segregación para variables continuas: <i>rank-order segregation index</i>	18
2.3 Segregación en el Sistema Escolar Chileno	19
3. DATOS	24
3.1 Fuentes Principales	24
3.2 Fuentes Auxiliares	25
3.3 Georeferenciación y Distancia	26
3.4 Variables a Nivel Individual	27
3.5 Variables a Nivel Colegio	29
4. CÁLCULOS DE SEGREGACIÓN	34
4.1 Descomposición <i>between</i> y <i>within</i>	35

4.1.1	Índice para variables dicotómicas	36
4.1.2	Índice para variables continuas	40
5.	MODELACIÓN Y ESTIMACIÓN	43
5.1	Modelación	43
5.1.1	Teoría de utilidad aleatoria	45
5.1.2	Logit de clase latente	47
5.2	Estimación	48
6.	DISPONIBILIDAD Y CONSIDERACIÓN DE ALTERNATIVAS	50
6.1	Construcción del Conjunto de Alternativas	53
6.1.1	Criterios de reducción de alternativas	53
6.2	Conjunto Final de Alternativas	55
7.	RESULTADOS MODELO DE ELECCIÓN	59
7.1	Clases	60
7.2	Preferencias (según clase)	60
7.3	Efectos Marginales	62
7.4	Reflexiones a partir de los resultados	65
8.	EJERCICIOS	72
8.1	Diseño	72
8.2	Resultados	73
8.2.1	Índice de segregación sin grupo “vulnerable”	76
9.	CONCLUSIÓN	78
	BIBLIOGRAFÍA	81
	ANEXOS	94
A.	BASE DE DATOS	95
A.1	Área geográfica definida para el trabajo	95
B.	INGRESOS IMPUTADOS	98

C. ANÁLISIS COMUNAL	100
C.1 Segregación	103
D. ÍNDICE DE SEGREGACIÓN PARA VARIABLE CONTINUA	108
D.1 Propiedades	108
D.2 Construcción	108
D.3 Descomposición	110
E. DEFINICIONES DE DISPONIBILIDAD	112
E.1 Criterios para Reducir Alternativas	112
E.1.1 Distancia	112
E.1.2 Ingreso	112
E.1.3 <i>Timing</i> de elección	114
E.1.4 Notas	115
E.1.5 Co-educacional	116
E.2 Disponibilidad Geográfica	118
E.3 Conjunto de Alternativas Aleatorio	119
F. SELECCIÓN ACADÉMICA	121
F.1 Movilidad	121
F.2 Perfil Académico de Nuevos Estudiantes	123
F.3 Colegios que Seleccionan	127
G. ESTIMACIÓN	129
G.1 Primera Etapa: <i>EM algorithm</i>	129
G.2 Segunda Etapa: Optimización Directa	130
G.2.1 Condiciones de primer orden	131
G.2.2 Condiciones de segundo orden	132
G.3 Tercera Etapa	133
G.4 Errores estándar	135
G.4.1 Otros errores estándares	135

G.5	Efectos Marginales	136
G.5.1	Variables continuas	136
G.5.2	Variables binarias	137
G.6	Ajuste	137
H.	GRÁFICOS DE EFECTOS MARGINALES	140
I.	EJERCICIOS	143
J.	REGLAS DE ASIGNACIÓN DE COLEGIO SEGÚN PREDICCIÓN	148

ÍNDICE DE FIGURAS

1.1.	Matrícula absoluta educación media 1995–2012.	7
1.2.	Participación relativa según dependencia en Gran Santiago 1995–2012. . . .	7
1.3.	SIMCE promedio vs ingreso per cápita promedio	8
4.1.	Índices de segregación para distintos percentiles de vulnerabilidad.	34
4.2.	Descomposición de H para distintos percentiles de vulnerabilidad	39
4.3.	Otras descomposiciones de H para distintos percentiles de vulnerabilidad . .	39
6.1.	Ubicación geográfica de los colegios según dependencia	57
6.2.	Ubicación geográfica de los individuos según decil de ingreso	58
7.1.	Medianas de tasas marginales de sustitución según decil de ingreso.	64
8.1.	Índices de segregación para distintos percentiles de vulnerabilidad.	74
8.2.	Descomposición de la segregación según diferentes índices	75
A.1.	Oferta de colegios enseñanza media en la Región Metropolitana	96
A.2.	Observaciones en Región Metropolitana	97
B.1.	Comparación de distribución geográfica del ingreso per cápita	99
C.1.	Cantidad de colegios y SIMCE promedio por comuna	101
C.2.	Relación entre cantidad de alumnos que emigran e inmigran por comuna. . .	101
C.3.	Correlaciones de variables comunales	103
C.4.	División del Gran Santiago en cuadrados	105
C.5.	Medición de segregación urbana según tamaño de la unidad	107
E.1.	Ingresos per cápita de familias por colegio	114
E.2.	Ejemplo de restricción de disponibilidad por nota	117

F.1.	Distribución de notas en colegios con mejor SIMCE con cobro $\leq$ \$7.500 . . .	124
F.2.	Distribución de notas en colegios con peor SIMCE con cobro $\leq$ \$7.500 . . .	125
F.3.	Distribución de notas en colegios con mejor SIMCE con cobro $\leq$ \$17.500 . . .	125
F.4.	Distribución de notas en colegios con peor SIMCE con cobro $\leq$ \$17.500 . . .	125
F.5.	Distribución de notas en colegios con mejor SIMCE con cobro $\leq$ \$60.000 . . .	126
F.6.	Distribución de notas en colegios con peor SIMCE con cobro $\leq$ \$60.000 . . .	126
F.7.	Distribución de notas en colegios con mejor SIMCE con cobro $>$ \$60.000 . . .	126
F.8.	Distribución de notas en colegios con peor SIMCE con cobro $>$ \$60.000 . . .	127
H.1.	Histogramas de efectos marginales a nivel de observación	141
H.2.	Efectos marginales según decil de Ingreso	142

ÍNDICE DE TABLAS

3.1.	Cálculos en base a las distancias a los colegios	27
3.2.	Medias de variables nivel individual por decil de ingreso	30
3.3.	Medias de variables nivel colegio por dependencia	33
4.1.	Descomposición del índice de teoría de la información.	37
4.2.	Descomposición del <i>rank-order information theory index</i>	40
6.1.	Cantidad de colegios disponibles promedio según decil y tipo de restricción .	55
7.1.	Características promedio de individuos según clase	61
7.2.	Tabla de parámetros modelo Logit Clase Latente, A	67
7.3.	Tabla de parámetros modelo Logit Clase Latente, B	68
7.4.	Tabla de parámetros modelo Logit Clase Latente, C	69
7.5.	Tabla de parámetros modelo Logit Clase Latente, D	70
7.6.	Efectos marginales de cambio atributos de colegio por decil de ingreso	71
8.1.	Resumen resultados ejercicios según diferentes índices para variable dicotómica	75
8.2.	Resumen resultados ejercicios según diferentes índices para variable continua	77
C.1.	Datos de matrícula y colegios a nivel comunal.	104
C.2.	Descomposición de segregación escolar y urbana a nivel comunal.	106
E.1.	Colegios disponibles en radio según decil, sin disponibilidad restringida . . .	119
E.2.	Colegios disponibles en radio según decil, con disponibilidad restringida . . .	119
F.1.	Flujo de cambios de colegio entre dependencias	122
F.2.	Cambios de colegio según decil de ingreso per cápita	123
F.3.	Colegios que seleccionan por notas y relación con SIMCE	128

G.1. Estadísticos de ajuste del modelo	138
G.2. Dependencia de la predicción vs dependencia de elección real	139
G.3. Correcta predicción entre tres primeras preferencias	139

RESUMEN

Existe hoy un diagnóstico bastante compartido: el nivel de segregación escolar en Chile es alto. Las razones que la causan no son consensuadas, y menos de la magnitud de éstas. En una aproximación empírica al problema, se estimó un modelo de elección de colegio para la educación media de Santiago. Se utilizó una estructura de elección discreta flexible (*latent class logit*), que permitió acomodar preferencias acorde a las características de las familias, y que consideró restricciones en la disponibilidad y selección académica. Se encontró que los atributos más relevantes en la elección de colegio (precio, distancia, SIMCE y tipo de enseñanza) eran valorados de forma bastante heterogénea dentro de la población. Familias con más recursos ponen mayor énfasis en el rendimiento, les es menos costoso viajar, evitan colegios con formación técnico profesional, y en los extremos actúan como si valoraran de forma positiva el cobro. Utilizando el modelo para predecir la elección de colegio, se calculó la segregación hipotética en diferentes contrafactuales, basados en un equilibrio parcial. Los principales resultados indican que el financiamiento compartido es responsable de entre un 10 % y 23 % de la segregación escolar, mientras que cuan segregada es la ciudad explica entre un 8 % y un 13 %, dependiendo del índice que se considere. Más importante es el efecto de la distinta valoración de los atributos según nivel de ingreso, al que se le atribuye entre un 41 % y 46 % del total de la segregación. Decisiones de las familias, y la gran variabilidad en preferencias, son muy relevantes en la configuración de la segregación. Políticas orientadas únicamente a la oferta son insuficientes si se busca mejorar el nivel de integración.

Palabras Claves: Elección de colegio, logit de clase latente, segregación escolar.

ABSTRACT

School segregation in Chile is very high, however, there is little clarity on the factors behind it and their relative magnitude. Mostly, it has been attributed to school choice and public policy. In principle, parents have an unrestricted choice of schools yet, a considerable number of State-funded schools can charge parents a co-payment fee. We modeled and estimate secondary school choice in Santiago de Chile using a flexible structure (latent class logit) which allows us to differentiate preferences according to each family's characteristics, and which considers restrictions based on the availability of schools and academic selection. The most relevant aspects affecting choice (cost, distance, quality, and curricula) are valued heterogeneously amongst the population. Comparatively, wealthier families particularly value school academic performance, they have less consideration on travel costs and school fees, and they avoid vocational schools. Based on the estimated model, and using different counterfactuals and indices, we concluded that 10 % to 23 % of the total segregation can be explained by the fee that some subsidized schools charge, 8 % to 13 % by residential segregation, and 41 % to 46 % can be attributed to the variability in the families' preferences.

Keywords: School choice, latent class logit, vouchers, school segregation.

1. INTRODUCCIÓN

El mes de marzo del 2011 marcó el inicio de un movimiento social liderado por estudiantes, con un gran respaldo ciudadano, cuya demanda principal era una educación “pública, gratuita, y de calidad”; estableciendo el tema como prioridad en la discusión pública. Diagnósticos nacionales e internacionales han mostrado las grandes brechas que existen en resultados educacionales para Chile dependiendo del nivel socioeconómico que domina los colegios (Heyneman, 2004; Hsieh y Urquiola, 2006; Mizala, González, y Romaguera, 2004). A esto se suman los altos niveles de segregación según ingreso, que dejan a Chile en los primeros lugares a nivel mundial (Willms, 2010), discutiendo fuertemente la capacidad del sistema educacional como promotor de la integración social (Valenzuela, Bellei, y de los Ríos, 2008).

En los últimos 20 años, el PIB ha crecido un 150 %¹, y el gasto público en educación escolar al menos se ha cuadruplicado²; sin embargo los resultados en desempeño escolar aún se alejan bastante del promedio de la OECD³. Si bien esfuerzos como los aumentos en la subvención, la jornada escolar completa, la ley del 15 %, la ley de subvención preferencial, la prohibición de selección hasta sexto y la nueva institucionalidad⁴ han estado enfocados en mejorar déficits del sistema⁵, y asegurar prácticamente total cobertura, la estructura impuesta en 1981 no se ha modificado en sus bases.

¹En base a cambio 1992-2012. Series Históricas CENDA-UC, www.cendachile.cl.

²Subió desde los 1.012.799 MM\$ gastados en 1995, a 4.284.682 MM\$ en el 2013 (Libertad y Desarrollo, 2012, MM\$: millones de pesos chilenos.).

³Pese a esto, Chile obtuvo el mejor resultado promedio de latinoamérica en las pruebas PISA 2009 y 2012, al mismo tiempo que se mantiene en los puestos más altos en correlación nivel socioeconómico del colegio con nivel de rendimiento (OECD, 2010, 2013b).

⁴Nuevo Consejo Nacional de Educación, Superintendencia de Educación Escolar y de la Agencia de Calidad de la Educación, entre otros.

⁵Ver Arellano (2000) para programas específicos, y contexto financiero, de la vuelta a la democracia.

Las elecciones presidenciales y parlamentarias de fines del 2013 han reafirmado la posición central que ocupa la educación dentro del debate nacional, ya sea por su importancia en la construcción de una sociedad democrática, su relación con la productividad del país, o su relevancia sobre el justo derecho en que todos los ciudadanos tengan posibilidades de realización⁶. El diagnóstico general indica que es necesario un cambio que permita mayor igualdad, integración y calidad, y al parecer el país está en condiciones de hacerlo.

Surge entonces una pregunta fundamental, dentro de muchas otras, que permita orientar políticas eficientes hacia un sistema más justo: ¿a qué, y en qué medida, se debe la segregación del sistema educativo? La investigación actual destaca el financiamiento compartido, la segregación urbana, la capacidad de los colegios de seleccionar a sus alumnos y las preferencias de los padres como las principales razones⁷ que juegan en contra de la integración; pero pocos esfuerzos se han dedicado a cuantificar los efectos.

El presente estudio tiene la ambición de orientar la respuesta a esta última pregunta. Utilizando un modelo de elección de colegio, que sigue muy de cerca el trabajo de Gallego y Hernando (2008b), busca cuantificar el efecto aislado de las razones de la segregación para la educación media del Gran Santiago. La estrategia utilizada consiste en comparar la segregación del sistema actual con la que habría en ciertos escenarios ficticios. Estos últimos intentan imitar un sistema educacional carente de ciertos atributos que obstaculizarían la integración, en los que las familias “eligen” colegio siguiendo las preferencias estimadas, para luego calcular la hipotética segregación.

El aporte viene dado desde la innovación en la modelación: al incorporar una especificación más flexible permite identificar de manera menos restringida las diferencias en

⁶Este estudio no pretende ser una reflexión sobre los problemas y consecuencias de la segregación escolar en sociedad, para el propósito revisar por ejemplo Briebe (2013); García-Huidobro (2007); Larrañaga y Sanhueza (2007); Valenzuela, Bellei, y de los Ríos (2010).

⁷Todas enmarcadas bajo un sistema de subvención a la demanda (*vouchers*) y “libre elección”. Ver más desarrollo de éstas en Hernando y Niklitschek (2013b).

preferencia por atributos de los colegios, según las características de cada familia. Otra novedad es que restringe la disponibilidad de alternativas para cada individuo, lo que busca simular una elección más cercana a la real, considerando restricciones por nivel de ingreso, distancia y selección académica. Se destaca también la mayor precisión en la estimación de la importancia de la distancia en la elección, pues se utiliza una base con georeferencia de los hogares. Además, es el primer estudio de elección de colegios para Santiago basado en un modelo de elección discreta enfocado en la educación media⁸, que considera los establecimientos particulares sin subvención estatal. Por último, se calculan por primera vez para el contexto escolar chileno índices de segregación para variables continuas⁹, permitiendo una mirada del nivel de integración más general, que evita la definición arbitraria de un grupo “vulnerable”.

El trabajo continúa con una breve descripción del sistema de educación chileno, para introducir el contexto en que se realiza el estudio. En el capítulo 2 se presenta una resumida revisión bibliográfica de tres temas relevantes para el estudio: la elección de colegios, con especial énfasis en los estudios realizados en Chile; metodologías de medición de segregación; y en tercer lugar, estudios anteriores específicos a la estratificación del sistema escolar en Chile. El capítulo 3 se refiere a la fuente y descripción de los datos utilizados, pasando a cálculos de segregación en base a éstos en el capítulo 4. La modelación y estimación se desarrolla en el capítulo 5, seguido por el 6 que detalla el problema y la propuesta respecto a la conjunto de alternativas que consideran las familias. La revisión y discusión de los resultados obtenidos del modelo de elección se detallan en la sección 7, que continúa con el capítulo 8, dedicado a los ejercicios de equilibrio parcial para cuantificar las causas de la segregación. Finalmente se desarrollan las conclusiones del trabajo en el capítulo 9.

⁸Chumacero, Gómez, y Paredes (2011) incluyen educación básica y media (y particulares) de forma conjunta en un modelo de elección binaria de colegio más cercano.

⁹Paredes, Volante, Zubizarreta, y Opazo (2013) miden la segregación considerando una variable en categorías con una metodología similar.

1.1 Sistema educacional chileno

En 1981 se lleva a cabo una gran reforma¹⁰, vigente en gran parte hasta el día de hoy, que traspasó la administración de los colegios públicos desde el gobierno central a las municipalidades. Un pilar importante fue el cambio del esquema de financiamiento hacia una subvención proporcional a la asistencia, sin discriminar entre colegios de propiedad pública o privada¹¹, fomentando la entrada de particulares al sistema escolar con aporte estatal¹².

De esta manera, se les entregaba la posibilidad a los apoderados de elegir libremente entre colegios públicos o privados, siendo portadores del *voucher* de subvención. El diseño, entre otras cosas, buscaba generar los incentivos para que la competencia entre los establecimientos por atraer más alumnos, y recibir más recursos, se traduzca en mejoras en calidad (Friedman, 1962). A pesar de contar con la misma subvención por estudiante, el sistema municipal mantuvo algunas exclusividades: la posibilidad de recibir aportes directos de la municipalidad u otras entidades gubernamentales¹³, y mayores rigideces en temas relacionados a los contratos de la planta docente y selección de alumnos¹⁴.

En 1988 se crea la ley 18.768 que permite el cobro de mensualidad a los padres, no teniendo gran impacto debido al diseño de los descuentos sobre la subvención, poco

¹⁰Impuesta en dictadura, y de las más importante desde 1860: el inicio oficial de la educación pública gratuita (financiada íntegramente por el estado).

¹¹Plana por alumno, mayor para educación media y para zonas rurales. Hasta 1989 con reajuste anual menor al IPC.

¹²Antes de la reforma existían algunos de estos colegios, pero recibían menos de la mitad de los fondos de los establecimientos públicos; en general estaban vinculados a fundaciones. Esto generó que pasaran de representar un 11 % de la matrícula en 1981, a un 51 % el 2012. La participación de los privados sin aporte estatal ha variado entre al 6 % y 9 % en todo el periodo.

¹³Con datos del 2002, Gallego (2006) estima que un 30 % de los gastos de los colegios se realizaban con fuentes diferentes a la subvención por alumno.

¹⁴Establecimientos municipales no pueden seleccionar a menos que tengan sobre demanda (tampoco expulsar un alumno, solo “transferirlo”); mientras que los particulares subvencionados, hasta el 2009, podían utilizar métodos de selección.

conveniente para los colegios. La vuelta a la democracia en 1990 trae consigo numerosos programas para enfrentar déficits en calidad y equidad, basados fundamentalmente en aumento de recursos, manteniendo casi intacta la estructura heredada.

A partir de 1993 entra en vigencia una reglamentación que cambió el diseño de los descuentos a la subvención (ley 19.247), logrando que el cobro a los apoderados significara efectivamente mayores recursos para los establecimientos, creando lo que hoy se conoce como el “financiamiento compartido”. El descuento es proporcional (escalonado) al monto cobrado, siendo el tope 4 USE¹⁵. La normativa no permite cobrar en colegios municipales de básica, y en media solo previo a la aprobación de la mayoría de los apoderados. Además, exige que entre un 5 % y 10 %, dependiendo del monto de cobro, de los ingresos aportados por los padres se destinen a becas.

En 1997 se inicia la implementación de la jornada escolar completa¹⁶, que implicó un 27 % y 17 % más de horas pedagógicas semanales para básica y media respectivamente, y horas explícitas para trabajo docente fuera del aula. Lo anterior se tradujo en un aumento de un tercio de la subvención por alumno, y fuerte inversión en infraestructura¹⁷.

El 2006 se obliga a los colegios con aporte estatal a que un 15 % de su matrícula corresponda a estudiantes vulnerables¹⁸, a los cuales no se les puede cobrar, a excepción de que no tengan demanda por estos cupos¹⁹. El 2008 se estableció la subvención escolar preferencial, ley SEP, que permite a los colegios que voluntariamente se acojan recibir una

¹⁵Mensualidades bajo 0,5 USE no implican menor subvención. El máximo descuento es de un 35 %; por lo que, con un monto promedio de subvención de 2,5 USE (Paredes et al., 2013), un colegio que cobra el máximo recibe 1,625 USE del estado. Esto se traduce en un ingreso total por alumno un 125 % superior a un establecimiento que no cobra. 1 USE: \$20.055 el 2013, \$13.504 el 2006.

¹⁶Uno de los cuatro pilares de la reforma escolar impulsada en el gobierno de Eduardo Frei Ruiz-Tagle.

¹⁷La extensión de jornada implicaba que ya no se podía atender a dos turnos (mañana y tarde) de estudiantes en un establecimiento, lo que permitía también su uso posterior para actividades extra-programáticas. Solo en 1999, uno de los años de implementación, se invirtieron cerca de 100.000 MM\$ en infraestructura.

¹⁸Según índice construido por la Junta Nacional de Auxilio Escolar y Becas (JUNAEB).

¹⁹Decreto 196 o “Ley del 15 %”. Rojas (2009); Treviño, Salazar, y Donoso (2011) plantean dudas sobre el conocimiento y efectividad de esta normativa.

mayor subvención por cada alumno prioritario²⁰, con la obligación de eliminar procesos de selección mediante entrevistas o test y el cobro a los beneficiados. Por último, el 2009 la ley general de educación prohíbe la selección en todo establecimiento con financiamiento estatal entre pre-kinder y sexto año básico²¹.

Hoy existen básicamente tres tipos de establecimientos escolares en Chile: colegios municipales, de propiedad y financiamiento público; colegios particulares con financiamiento estatal, desde ahora “colegios subvencionados”; y colegios particulares sin aportes del gobierno, en adelante “colegios particulares”. El sistema municipal se puede dividir entre los establecimientos administrados directamente por la municipalidad; los manejados por una corporación municipal, dedicada exclusivamente a educación; o los de administración delegada, que se les entrega la gestión a una institución privada²². En el caso de los subvencionados se podrían separar en colegios con y sin fines de lucro²³, colegios con copago y sin copago, o religiosos y no religiosos.

La Figura 1.1 representa la matrícula de la educación media para el periodo 1995–2012, en miles de alumnos²⁴. Se aprecia que el 2006 fue el año con mayor cantidad de

²⁰Cerca de un 40 % población escolar cabe dentro de esta categoría, definida por el Ministerio de Educación y basada en dimensiones socioeconómicas (no coincide necesariamente con IVE construido por JUNAEB). Se entregan además recursos adicionales según concentración de alumnos prioritarios. Ver Mizala y Torche (2013) para mayor detalle.

²¹Elacqua y Santos (2013) discuten la ambigüedad de esta ley, y Carrasco, Bogolasky, Flores, Gutiérrez, y San Martín (2014) realizan un estudio sobre las prácticas de selección, dejando en evidencia su uso a pesar de las normativas.

²²Estrictamente son propiedad del estado (ley 3166), no de los municipios, pero en este estudio se incluyen dentro del grupo municipal (32 colegios).

²³De acuerdo con Elacqua y Santos (2013), al día de hoy es difícil de confirmar la condición de lucro o no lucro, pues si bien los colegios que pertenecen a fundaciones no pueden ser negocio, no necesariamente re invierten todos los excedentes. A la vez, existen instituciones que no buscan lucrar, pero por facilidades jurídicas prefieren tener su colegio al alero de una sociedad en vez de una fundación.

²⁴Respecto a la cobertura, en 1970 un 20 % de jóvenes entre 15 y 18 años cursaba enseñanza media (Arellano, 2000); hoy se espera que un 83 % de la población joven la complete, ligeramente superior a la tasa promedio en países de la OECD (OECD, 2013a).

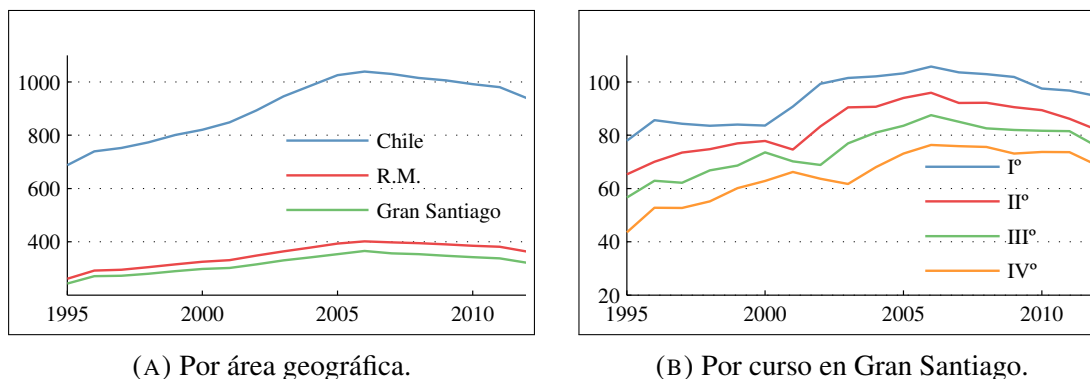


FIGURA 1.1. Matrícula absoluta educación media 1995–2012.

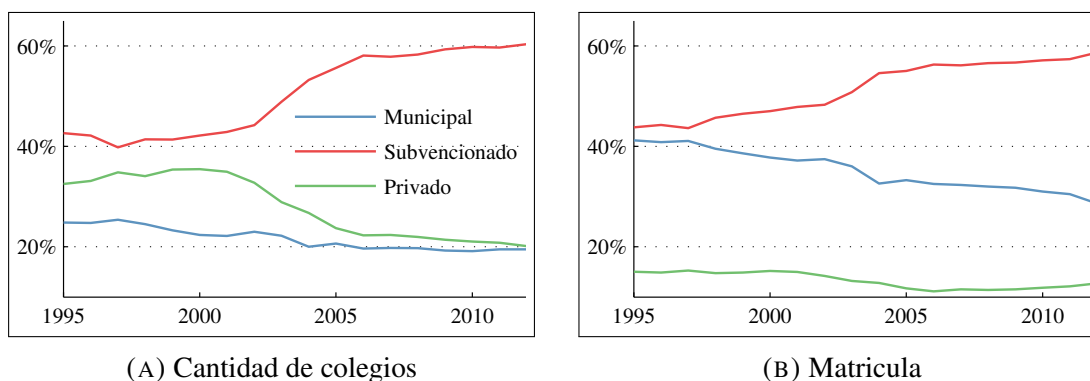


FIGURA 1.2. Participación relativa según dependencia en Gran Santiago 1995–2012.

alumnos en educación media, y que la transferencia año a año de alumnos no ha sido completa (importante deserción o repitencia). Al 2012 había 321.112 estudiantes en educación media en el Gran Santiago, repartidos en 912 colegios: un 10 % ofreciendo un currículo técnico profesional, un 70 % solo científico-humanista, y el resto mixto (polivalente). Se observa en la Figura 1.2 que los colegios subvencionados han aumentado su presencia, representando hoy un 60 % de los establecimientos y un 59 % de la matrícula. Estas mismas cifras para los de dependencia municipal son 19 % y 29 %, lo que ha significado un importante retroceso respecto a los años anteriores. Una última realidad se observa en la Figura 1.3, que relaciona resultado SIMCE con nivel de ingreso promedio de las familias del colegio; con datos del año 2006 y solo establecimientos que imparten educación media

del Gran Santiago (los colores representan el monto del copago, y la forma la dependencia). Se observa una correlación positiva, la que según informes de la prueba PISA es de las más fuertes a nivel mundial.

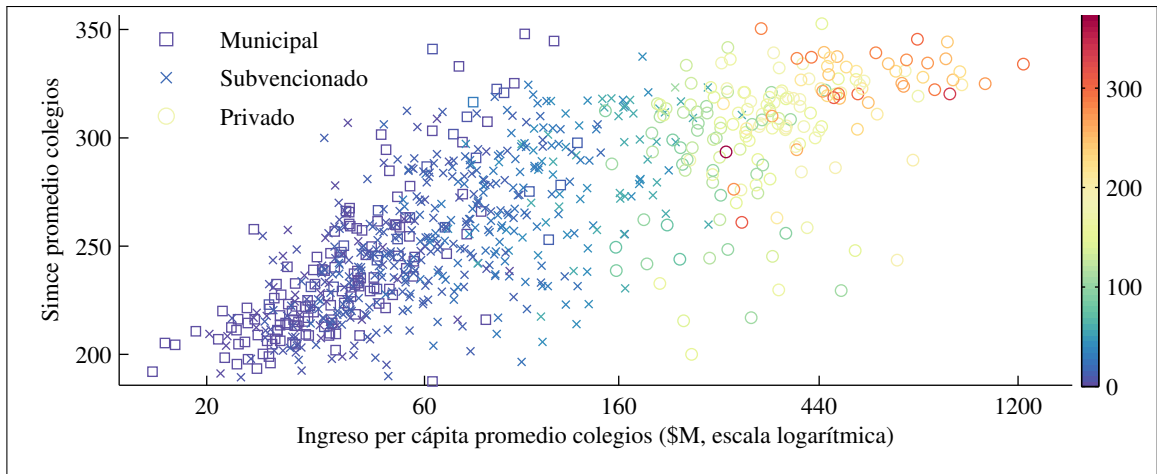


FIGURA 1.3. SIMCE promedio vs ingreso per cápita promedio (color representa mensualidad)

2. BREVE REVISIÓN BIBLIOGRÁFICA

2.1 Preferencias en Elección de Colegios

La implementación de políticas educacionales orientadas a una mayor participación de los padres en la elección de colegio, ha traído consigo bastantes trabajos orientados a la medición su impacto en diferentes dimensiones, pero pocos centrados directamente en el estudio de las preferencias (Burgess, Greaves, Vignoles, y Wilson, 2010). Ladd (2002); Levin (1998); Neal (2002) entre otros, recorren experiencias globales de sistemas con libre elección, encontrando diferentes resultados referidos a desempeño y equidad.

Ciertos críticos advierten que factores como las diferencias en capacidad de recopilación e interpretación de información, o los elementos que restringen la libertad (como la segregación urbana, el cobro o la capacidad de selección), generan inevitablemente un sistema estratificado por nivel socioeconómico¹. Por otro lado, quienes están a favor destacan el beneficio de poder elegir un proyecto educativo acorde a las preferencias de cada familia, y las posibilidades de los más pobres de optar por alternativas distintas a las de su entorno social o geográfico. Afirman que la competencia se traduce en mejores estándares, y que colegios que no entreguen buena educación fracasarán por no tener demanda, defendiendo las ventajas del mercado sobre un sistema centralizado².

En un trabajo orientado al proceso de elección, Hastings y Weinstein (2008) sugieren que el menor interés que ponen las familias en resultados académicos podría deberse a que, a pesar de valorar altamente lo académico, no pueden expresar dicha preferencia por los costos o restricciones de acceso e interpretación de la información, eligiendo finalmente colegios en base a características más evidentes como cercanía. Hastings, Van Weelden, y Weinstein (2007) encuentran gran heterogeneidad en preferencias analizando las tres

¹En extremo “...the implementation of market reforms in education is essentially a class strategy which has as one of its major effects the reproduction of relative social class (and ethnic) advantages and disadvantages.” (Ball, 1993).

²Gallego y Sapelli (2007) revisan en base a evidencia empírica algunos argumentos a favor y en contra para el caso chileno

primeras opciones escogidas por cada familia en el proceso de elección, y concluyen que la entrega de información aumenta la probabilidad de elegir un colegio con mejor rendimiento académico. La evidencia internacional destaca el compromiso que enfrentan las familias entre distancia y resultado en pruebas estandarizadas, y la menor intensidad en las preferencias por esto último de los más pobres³.

2.1.1 El caso chileno

Si bien el desempeño del modelo educacional chileno actual depende de la calidad de las elecciones de los padres, y la competencia que éstas generarían (Friedman, 1962), la investigación referida a elección de colegio (*school choice*) no es extensa. En esta sección se revisan los principales trabajos nacionales del área.

Sapelli y Torche (2002) realizan el primer estudio relacionado con modelar el proceso de elección de colegio en Chile. Los autores buscan responder qué determina la elección entre los tres tipos de colegios existentes, usando datos de educación básica provenientes de la encuesta CASEN 1996. Destacan la importancia de la movilidad en el sistema de subvención a la demanda: las restricciones atentan contra la eficiencia y limitan la elección. Concluyen que padres de mayor ingreso y educación tienen una mayor probabilidad de elegir un colegio subvencionado⁴.

Elacqua y Martínez (2010) trabajan con entrevistas a apoderados de primeros básicos de la Región Metropolitana, realizadas los años 2003 y 2009⁵, comparando las fuentes de información utilizadas para fundar la elección, la disposición al viaje y la satisfacción con

³Ver Elacqua (2012); Elacqua y Martínez (2010); Raczynski, Salinas, De la Fuente, Hernández, y Lattz (2010) para una revisión de la literatura internacional de elección de colegio en contexto de subvención a la demanda; o Gorard (1999) o Belfield y Levin (2002) para una revisión en el contexto del sistema británico y estadounidense respectivamente.

⁴Resultado de alcance limitado dada la gran heterogeneidad intra subsistemas, y el bajo conocimiento de la propiedad (si el colegio es público o privado) declarado por los padres de recursos medios y bajos (Raczynski et al., 2010).

⁵Entre medio se inicia la implementación la subvención escolar preferencial, y algunos programas de entrega de información para la elección, entre otras cosas.

el colegio. Para ambos cortes la mayor fuente información reportada por los padres son sus redes sociales, seguido por la que entrega el personal del colegio, *marketing*, Internet, el gobierno y, en último lugar, la obtenida de medios de comunicación. Encuentran un aumento en la cantidad de fuentes consultadas entre los dos periodos, siendo Internet la de mayor crecimiento, y una mayor disposición a asistir a colegios más alejados.

Raczynski et al. (2010) realizan entrevistas a estudiantes y apoderados de estratos medios y bajos⁶ con el objetivo de indagar en la valoración de los establecimientos según sector y conocer el proceso de elección de colegio. Los resultados sugieren que los padres diferencian los colegios por el monto del cobro más que por la dependencia, y buscan estar con pares similares, para evitar malas influencias y un clima no apropiado para el estudio. Lo anterior lleva a que por lo general valoren positivamente la selección. Las características que los apoderados relacionan con calidad son disciplina, preocupación por el estudiante, aprendizaje y formación valórica; y los dos factores más relevantes son costo y distancia. Los autores indican que la información más importante para la elección de colegio proviene de las redes de los padres (o estudiante), y la que ellos mismo generan a partir de observaciones. Concluyen que los criterios de elección son heterogéneos, dependiendo del nivel de ingreso, cultural y de expectativas (motivación).

En el estudio que inspiró este trabajo, Gallego y Hernando (2008b) modelan la elección de colegios mediante un modelo logit multinomial, utilizando la especificación de Berry, Levinsohn, y Pakes (1995)⁷. Utilizando observaciones del SIMCE de cuartos básicos del año 2002, estiman parámetros en función (lineal) de los atributos de las familias, considerando así heterogeneidad en las preferencias. Dentro de los resultados destacan que los padres ponen gran énfasis en calidad y distancia, creciendo la importancia de la

⁶Entre entrevistas personales y grupales, conversan con 124 apoderados y 24 alumnos, en las comunas de Puente Alto, Osorno y Coquimbo

⁷El modelo incluye un efecto fijo por colegio no observable para el investigador, común a todos los individuos, y corrige la posible endogeneidad debido a la correlación entre calidad no observable y precio o rendimiento.

primera y disminuyendo la segunda a medida que aumenta el ingreso familiar. Otros atributos que encuentran relevantes en la elección son el costo, la disciplina, la formación valórica y la composición de género. A partir de estimaciones independientes, con grupos de colegios con distintas exigencias en el proceso de admisión, los autores sugieren que es más relevante el fenómeno de auto-segregación que la selección realizada por los establecimientos sobre la segregación del sistema escolar.

En una extensión del trabajo anterior, Gallego y Hernando (2008a) cuantifican el “valor” de la opción de elegir a través de construcción de contrafactuales⁸. Calculan que la “libre elección” tiene un beneficio neto positivo en bienestar (variación compensatoria), en comparación a un escenario de distribución aleatoria⁹. Muestran que la repartición de esta ganancia no es pareja, sino que está concentrada en el 40 % de la población con mayores ingresos, no siendo beneficiados el resto de los individuos con una política educacional de “libre elección”. Por otro lado, encuentran que la movilidad es mayormente benéfica para las familias con menos recursos, si se compara el bienestar entre el modelo actual y uno con elección limitada a la comuna de residencia.

Siguiendo una línea similar, Chumacero et al. (2011) estiman un modelo de elección de colegio más cercano¹⁰ utilizando datos georeferenciados de la encuesta CASEN 2003, correspondientes a observaciones de educación básica y media. Concluyen que la distancia y la calidad son los atributos más relevantes en la elección del colegio más cercano.

Gallego, Lagos, y Stekel (2012) evalúan el impacto de entregar información sobre la “calidad” de los colegios en el barrio¹¹ en la elección de colegio para padres de bajos recursos mediante un experimento aleatorio. El trabajo incluye entrevistas ($N = 1.611$) que

⁸Escenarios que buscan representar lo que pasaría en alguna condición distinta a la realidad (por ejemplo la ausencia de una política o atributo que existe, o la presencia de uno que no existe).

⁹Asignando las observaciones únicamente a colegios dentro de la comuna de residencia.

¹⁰Por costos computacionales y facilidad en la interpretación de resultados, los autores definen el conjunto de alternativas de cada individuo como el colegio real elegido y el colegio más cercano, construyendo así una estructura de elección binaria que modelan con un modelo Probit.

¹¹También informan sobre la rentabilidad de un “buen colegio”.

indagan sobre el proceso de elección, encontrando que los dos atributos más relevantes son la cercanía y el resultado SIMCE (siendo el primero el más importante para el grupo que no recibe información, y el segundo para el que sí recibe). Concluyen que informar sobre la calidad de los colegios del sector aumenta el rendimiento, precio y distancia promedio del colegio elegido.

Flores y Carrasco (2013) realizan un análisis de entrevistas a padres con hijos entre 2 y 5 años en etapas previas al proceso de matrícula, con el fin de poder aislar las preferencias puras de los padres¹². Concluyen que cercanía, resultado SIMCE y bajo cobro son tres atributos que se asocian a la “popularidad” de un colegio (medida como la cantidad de padres que lo incluyen dentro de sus alternativas). Considerando la ubicación de los hogares, los autores sostienen que familias de mayor nivel socioeconómico toman en cuenta alternativas más distantes de sus residencias, y tienen acceso a colegios con mejor resultado SIMCE. Sostienen que esto último explicaría parcialmente la elección de colegios de peor rendimiento por parte de familias pobres, quienes sí valorarían el desempeño, pero cuya restricción de alternativas afectaría en la consecución de sus verdaderas preferencias.

Por último, Grau (2014) estima un modelo de elección de colegio dinámico, dentro un horizonte de ocho años (2003-2011), enfocado en educación básica de Santiago. La modelación permite que distintos individuos tengan diferentes percepciones de la calidad de un mismo colegio (debido a la confusión entre el valor agregado que produce el colegio y el nivel socioeconómico del alumnado), y restringe el conjunto de alternativas disponibles a las elecciones realizadas por padres de la misma comuna con educación similar. Además, toma en cuenta la selección escolar a través de una función construida con los requisitos del proceso de admisión reportados en el cuestionario para padres del SIMCE y las características de los estudiantes. El autor concluye que existe una gran diferencia entre la calidad percibida y la real (estimada), efecto de mayor magnitud en padres con menor educación, y que sobrestima más la calidad en colegios subvencionados y privados,

¹²De acuerdo a los autores, libres de efectos de restricción o selección, que están presentes si se toma en cuenta el colegio matriculado como la elección.

afectando negativamente la percepción sobre los colegios públicos. Por medio de contrafactuales, evalúa el impacto en la demanda de un cambio exógeno positivo en la calidad de colegios escogidos aleatoriamente. Encuentra un efecto moderado en la demanda cuando el aumento es anterior a la decisión del primer año (decisión de matrícula), e irrelevante cuando la decisión es en años posteriores (decisión de cambio).

2.2 Medición de Segregación

La aproximación al fenómeno de segregación es desde una de las cinco dimensiones estudiadas en Massey y Denton (1988): uniformidad (*evenness*)¹³. Siguiendo a los autores, esta dimensión se entiende, en su forma más general, como el grado en que cierto grupo está distribuido de forma pareja en todas las unidades. Lo anterior significa que para considerar nula la segregación, la proporción que representan los miembros del grupo de interés para el que se evalúa la segregación, tiene que ser igual en cada una de las unidades de análisis. En el contexto de este estudio, el grupo de interés está constituido por todas las observaciones con ingreso per cápita menor a cierto valor (en general el 30 % de las observaciones con menor ingreso¹⁴), a quienes se les llamará “vulnerable”. Las unidades de análisis son los colegios.

2.2.1 Índices de segregación para grupos dicotómicos

La necesidad de construir índices surge a partir de la imposibilidad de comparar niveles de segregación únicamente a través de curvas de Lorenz¹⁵. Existe bastante debate en la literatura sobre qué índice debiera usarse (Allen y Vignoles, 2007; Cortese, Falk, y Cohen, 1976; Duncan y Duncan, 1955; Gorard y Taylor, 2002; Hutchens, 2001; James y Taeuber, 1985; Massey y Denton, 1988), pues, además de variar en propiedades (Hutchens, 2004),

¹³Las otras cuatro que desarrollan los autores son exposición (*exposure*), concentración (*concentration*), centralización (*centralization*) y agrupamiento (*clustering*); las tres últimas referidas principalmente al problema de la segregación urbana.

¹⁴Pero varía según el análisis. Para todos los casos está ligado a algún percentil de la distribución de ingreso.

¹⁵Curvas que representan la distribución de la variable de interés en la unidades de análisis, permitiendo para algunos casos poder comprar el grado de segregación.

la elección implica algún juicio de valor relacionado con cómo se entiende la medición de segregación (Atkinson, 1970).

Definiendo J como el número total de colegios, t_j , p_j , T y P como el numero de individuos y proporción de “vulnerables” en colegio j y totales respectivamente, se enumeran a continuación los índices utilizados:

- Índice de Duncan¹⁶, representado por la letra “ D ” (ecuación 2.1, Duncan y Duncan, 1955).

$$D = \frac{1}{2} \sum_{j=1}^J \frac{t_j}{T} \frac{|P - p_j|}{P(1 - P)} \quad (2.1)$$

- Índice de teoría de la información o *information theory index*, representado por “ H ” (ecuación 2.2, Theil y Finizza, 1971)

$$H = \sum_{j=1}^J \frac{t_j}{T} \frac{P \log_2(P) + (1 - P) \log_2(1 - P) - (p_j \log_2(p_j) + (1 - p_j) \log_2(1 - p_j))}{P \log_2(P) + (1 - P) \log_2(1 - P)} \quad (2.2)$$

- Índice de tasa de varianza o *variance ratio index*¹⁷, representado por “ R ” (ecuación 2.3, James y Taeuber, 1985)

$$R = \sum_{j=1}^J \frac{t_j}{T} \frac{(P - p_j)^2}{P(1 - P)} \quad (2.3)$$

- Índice de raíz cuadrada o *square root index*, representado por “ S ” (ecuación 2.4, Hutchens, 2001).

$$S = \sum_{j=1}^J \frac{t_j}{T} \frac{\sqrt{P(1 - P)} - \sqrt{p_j(1 - p_j)}}{\sqrt{P(1 - P)}} \quad (2.4)$$

Todos los indicadores están acotados entre 0 y 1, tomando el valor mínimo cuando hay integración perfecta, es decir cuando todas las unidades tienen la misma proporción del

¹⁶También conocido como índice de disimilitud (*dissimilarity index*).

¹⁷Eta cuadrado en Duncan y Duncan (1955), S en Zoloth (1976) o tasa de correlación en Farley (1977).

grupo minoritario¹⁸, y máximo cuando existe segregación perfecta. Además, son simétricos, no importa cuál de los dos grupos que define la variable de interés¹⁹ se denomine el “vulnerable”, pues el resultado será el mismo.

La interpretación más clásica para el índice de Duncan es el porcentaje de individuos del grupo “vulnerable” que tendría que ser transferido a otros colegios para igualar la proporción de éstos en todas las unidades. Otra interpretación menos popular, y directa de la ecuación (2.1), es la diferencia absoluta promedio (ponderada) entre la proporción de “vulnerables” de las unidades y del sistema.

El índice de teoría de la información se basa en el concepto de entropía²⁰, y se interpreta como el promedio ponderado de la diferencia entre la entropía (grado de diversidad) de las unidades y el máximo de entropía alcanzable dada la proporción de “vulnerables” del sistema (que equivale a la entropía del sistema). Entonces, si en todas las unidades los “vulnerables” representan la misma proporción, el índice toma el valor de cero. Por el contrario, si no existe diversidad en las unidades (es decir, algunas se componen en su totalidad por “vulnerables”, y otras de solo no “vulnerables”), entonces el promedio de entropía de éstas será cero y el índice toma el valor máximo.

De manera análoga a la segunda interpretación del Duncan, el índice de tasa de varianza es la diferencia cuadrática promedio (ponderada) entre la proporción de “vulnerables”

¹⁸Diversos autores han criticado utilizar el valor cero como escenario base de no segregación (Åslund y Skans, 2009; Carrington y Troske, 1997; Cortese et al., 1976; Winship, 1977), proponiendo que, en la práctica, el escenario de segregación ideal debiera ser el resultado de una asignación aleatoria (índice de segregación potencialmente mayor a cero), más que el de distribución perfecta (valor de cero).

¹⁹Recordar que son índices basados en la distribución de cierto grupo caracterizado por una variable dicotómica, que divide las observaciones en dos grupos.

²⁰Entropía en el contexto de teoría de la información se puede interpretar como una medida de la impredecibilidad de cierto contenido de información (Shannon, 2001). Este concepto se ha utilizado para medir “entropía racial” (Theil y Finizza, 1971), y se asocia con una medida de diversidad (Reardon y Firebaugh, 2002). En su definición más tradicional es una función cóncava simétrica, fluctúa entre 0 y 1, toma el valor mínimo cuando el porcentaje de “vulnerables” es 0 % ó 100 %, y máximo cuando es 50 %.

de las unidades y del sistema; o la razón entre la varianza de las proporciones de “vulnerables” en las unidades y la varianza de una distribución binomial con media igual a la proporción de “vulnerables” del sistema.

Por último, el índice de raíz cuadrada surge como crítica a los ya mencionados (en especial al Duncan y al Gini), por no cumplir ciertas propiedades propuestas como requisito para un “buen índice” (Hutchens, 2001). Jenkins, Micklewright, y Schnepf (2008) muestran que este índice se puede interpretar como el promedio de la diferencia entre la media geométrica de las proporciones de grupos (para este caso “vulnerables” y no “vulnerables”) de cada unidad y la media que habría sin segregación.

Zoloth (1976) enfatiza la ventaja del uso de los índices H y R sobre D , dado que los primeros castigan de manera “más severa” unidades con mayor diferencia entre su proporción de “vulnerables” y la del sistema, mientras que D penaliza de forma lineal²¹. El autor argumenta que la segregación es más costosa socialmente (en términos marginales) cuando está más alejada de un escenario de integración, que cuando lo está menos, por lo que es deseable que el indicador refleje esta diferencia. Además, esta propiedad incentiva a hacer política donde los niveles de segregación son más altos, debido a que el retorno será más grande.

Reardon (2011) muestra empíricamente que H y S son más sensibles a la segregación de grupos extremos que R , por lo que propone que son de especial interés cuando es relevante la segregación de los muy pobres o muy ricos.

²¹Por ejemplo, dos unidades A y B de un sistema con varias otras unidades (por simplicidad todas con igual capacidad), con proporción minoritaria alejada en 0,1 a la proporción poblacional; tiene la misma segregación según D que un sistema en que A tiene una proporción igual a la del sistema y B alejado en 0,2. Mientras que calculando H , sería más alto para el segundo caso,; pues considera “más grave” que exista un colegio muy segregado, a dos con niveles de segregación de “la mitad”: el aporte marginal al índice de segregación es creciente al alejarse de la proporción del sistema.

2.2.2 Índices de segregación para variables continuas: *rank-order segregation index*

Reardon (2011) propone un índice para evaluar segregación cuando la variable de interés es continua, evitando así el proceso arbitrario de definir un límite para características no dicotómicas, y la pérdida de información asociada a dividir un grupo ordenado en dos categorías²². La interpretación más general de este indicador es en qué medida observaciones con distinto valor en la variable de interés (estudiantes con distinto ingreso familiar para el caso) están distribuidos de manera pareja en todas las unidades (colegios).

El índice solo se ve afectado frente a cambios en la variable de interés que modifiquen la posición relativa (o ránking) de una observación con respecto al resto. Esto implica que la medición se relaciona únicamente con la diferencia entre la distribución de la variable en las unidades y la distribución en la población, no con el nivel de desigualdad²³.

La ecuación (2.5) representa el indicador, donde J es el número de unidades, t_j el número de observaciones por unidad, T el total de observaciones. u y u_j son una medida de la dispersión (o distribución) de la variable para la que se está evaluando la segregación, en la población y unidad j respectivamente. La resta $u - u_j$ puede ser interpretada como el grado de diferencia entre la distribución de la variable de interés en la población y la distribución en una unidad j . Por lo tanto, un valor bajo en este índice de segregación significaría que las distribuciones de la variable de interés en las unidades son similares a las de la población: las observaciones están distribuidas de manera pareja en términos de la variable a considerar para la segregación.

$$\Lambda = \sum_{j=1}^J \frac{t_j}{T} \left(\frac{u - u_j}{u} \right) \quad (2.5)$$

²²Si la variable respecto a la que se quiere medir la segregación es continua, entonces el grupo es ordenado en el sentido que se podría hacer un ránking de las observaciones y construir un índice que considere toda la información, evitando la pérdida al transformar la variable continua en binaria.

²³De hecho, para la construcción del índice solo basta el percentil al que pertenece cada observación, y no la magnitud de la variable de interés.

Dependiendo de la definición que se escoja para u y u_j , el autor construye tres índices: *rank-order information theory index (R-O H)*, *rank-order variance ratio index (R-O R)* y *rank-order square root index (R-O S)*. Además, demuestra que no son más que un promedio ponderado de los índices teoría de la información, tasa de varianza y raíz cuadrada respectivamente, calculados para distintos límites de vulnerabilidad. En el anexo D se describen algunas de las propiedades y se explica con detalle la construcción del índice.

2.3 Segregación en el Sistema Escolar Chileno

La investigación nacional relacionada con segregación escolar ha estado en auge desde fines de la década pasada, existiendo una cantidad reducida de trabajos anteriores al 2008. En esta sección se resumen los principales resultados de investigación empírica en temas de segregación escolar del sistema chileno.

Con información entre 1982 y 1996, y analizando a nivel de mercados independientes²⁴, Hsieh y Urquiola (2006) encuentran correlación negativa entre crecimiento del sector privado subvencionado y nivel socioeconómico de las familias en colegios municipales, sin encontrar evidencia de que en mercados con mayor aumento de matrícula en colegios subvencionados haya mejorado el sistema²⁵. Sugieren que el principal resultado de los cambios de 1981 es una migración de la clase media desde la educación municipal hacia los colegios subvencionados (fenómeno que asocian a *cream-skimming*), y no una mejora en la calidad global²⁶.

²⁴Definidos como cada comuna, a excepción de Santiago en que se agrupan 50 comunas para constituir un solo mercado.

²⁵Evaluando específicamente resultados pruebas estandarizadas, tasa de repitencia y años de escolaridad.

²⁶Los autores no descartan que puede haber mejorado en dimensiones no consideradas, pero no por eso poco relevantes para las familias.

En un trabajo empírico enfocado en la segregación, Valenzuela et al. (2008) calculan un índice de Duncan²⁷ de 0,43 y 0,49, incluyendo todos los establecimientos urbanos de segundo medio de los años 2001 y 2006 respectivamente (0,31 y 0,37 para municipal; 0,40 y 0,48 para subvencionado; 0,65 y 0,91 para particular; y 0,44 y 0,50 para RM)²⁸. El estudio finaliza con un análisis multivariado en que encuentran correlación positiva entre segregación educacional comunal (medida con Duncan) y mayor participación de colegios subvencionados con copago.

Gallego y Hernando (2008a) calculan un índice Duncan de 0,39 utilizando datos de Santiago a nivel censal, correspondientes cuartos básicos del 2002. En el mismo trabajo, estiman por medio de un modelo de elección discreta (Gallego y Hernando, 2008b) el efecto de eliminar el copago, lo que resulta en un Duncan de 0,34. Mencionan que la disminución en segregación es “a lo más moderada”. Sugieren que la segregación se explicaría mayormente por factores relacionados con la demanda más que selección de parte de la oferta.

Mizala y Torche (2012), con datos de cuartos básicos del 2002 y octavos básicos del 2004, muestran que el sistema subvencionado, como un todo, atiende un público bastante heterogéneo en términos socioeconómicos, pero bastante homogéneo si se mira nivel de colegio. Argumentan que el diseño del sistema genera colegios especializados en diferentes mercados segmentados por nivel de ingreso. También encuentran fuerte correlación entre rendimiento y nivel socioeconómico del alumnado, efecto doble en magnitud en sector subvencionado que en el municipal. Su resultado más relevante es que, controlando por nivel socioeconómico, encuentran que el copago no está correlacionado con mejores resultados SIMCE.

²⁷Consideran el 30 % de menor nivel en un índice socioeconómico construido mediante componentes principales, en base a la educación de los padre e ingreso per cápita, datos obtenidos de la Encuesta Complementaria a las Familias del SIMCE. Recordar que un valor de “0” para el índice de Duncan significa integración perfecta y “1” lo contrario.

²⁸Utilizan también otros grupos “vulnerables” a partir del nivel educación de la madre, realizan lo mismo para 4tos y 8vos básicos y calculan un índice de aislamiento.

En un estudio que considera alumnos prioritarios según IVE²⁹ como el grupo “vulnerable”, Elacqua (2012) calcula un índice Duncan que varía entre 0,35 y 0,38 para el sistema con financiamiento estatal entre el 2000 y el 2006, y entre 0,23 y 0,28 y 0,42 y 0,45 para los subsistemas municipales y subvencionado respectivamente³⁰. Estiman una leve tendencia a la baja en la segregación con el aumento de los años.

Paredes et al. (2013) construyen índices de segregación para educación básica y media utilizando observaciones de todo Chile, basados en datos de SIMCE de los años 1998 a 2010. Considerando el grupo “vulnerable” como las familias con ingreso menor a \$200.000, calculan índices de Duncan para educación media que varían entre 0,46 y 0,51, sin aparente tendencia temporal (para el 2006 estiman la segregación del sector municipal en 0,37, subvencionado en 0,46, y particular en 0,76). Realizan el ejercicio de calcular segregación con subgrupos de colegios, quitando gradualmente los establecimientos con mayor cobro. Concluyen que gran parte de la segregación podría entenderse, más que por la presencia de colegios con copago, por el monto que se cobre. Por último, muestran que colegios con baja mensualidad presentan niveles de segregación incluso menores al sistema municipal.

La fundación Espacio Público lanzó a fines del 2013 una serie de estudios relacionados con segregación escolar. En uno de los trabajos, utilizando datos del SIMCE de cuartos básicos del 2010 del Gran Santiago, Elacqua y Santos (2013) muestran que los barrios son menos segregados que los colegios³¹, y que establecimientos municipales y subvencionados sin fines de lucro no representan la composición de sus barrios: en los

²⁹Índice de Vulnerabilidad Escolar, construido por la JUNAEB en base a encuestas a los apoderados hechas por los colegios. De los mismos cálculos del autor, cerca de un 40 % de la población escolar municipal es considerada como prioritaria, un 20 % de la perteneciente a colegios subvencionados y menos de 1 % de los asistentes a colegios particulares. Estos últimos datos son estables entre los años 2000 y 2006.

³⁰Solo considera los establecimientos en zonas urbanas para los cálculos, no detalla que cursos son parte de la muestra. Desagrega también con y sin copago, calculando índices entre 0,40 y 0,46 y 0,31 y 0,38 respectivamente.

³¹Distribuyendo las observaciones en el colegio más cercano, al parecer sin respetar la capacidad de los establecimientos.

primeros los más vulnerables están sobre-representados y en los segundos los con más recursos³². En otro esfuerzo, Valenzuela, Villalobos, y Gómez (2013) calculan índices de Duncan considerando 100 límites de vulnerabilidad, correspondientes a los percentiles de la distribución de ingreso y de resultado en SIMCE. Muestran que los grupos extremos son los más segregados³³, y que en general la segregación aumenta entre 1999 y 2011. Comparando la segregación de diferentes grupos, concluyen que en el sistema actual se poco probable la interacción entre miembros de los tres deciles de mayor ingreso y los grupos adyacentes, y menor aún la integración entre el 10 % más rico y la clase media y media-alta.

En resumen, no hay consenso de la evolución, tampoco se ha testado estadísticamente³⁴. Los niveles de segregación estimados son en general mayores para el sistema subvencionado que el municipal, y muy superiores para el privado. Además, son sensibles a qué índice se utilice para medir la segregación, a la definición del grupo vulnerable, así como a el área geográfica que se abarque o el nivel educacional que se utilice. Entonces, ¿es alta la segregación en el sistema escolar? Si se compara con niveles de otros países, sí (Willms, 2010); si se considera alta la segregación urbana, entonces la escolar lo es mucho más; y si se piensa que tener que trasladar a un porcentaje cercano al 40 % de la población escolar vulnerable para lograr integración es bastante, entonces el sistema está segregado. Al menos no hay evidencia de que presente niveles de hiper-segregación, de acuerdo a la definición de Glaeser y Vigdor (2003)³⁵.

En el capítulo 4 se realiza un análisis más de la segregación para la educación media del Gran Santiago. Se consideran varios niveles de “vulnerabilidad” y descomposición del

³²Se excluyen del estudio los colegios y alumnos del sistema privado.

³³Como ya se ha mencionado, la forma de “u” en los perfiles de segregación según percentil puede deberse en parte a un resultado mecánico de la medición, debido a la sensibilidad de los índices.

³⁴Mora y Ruiz-Castillo (2003) enumeraron trabajos enfocados en procedimientos que permiten testear cambios en segregación.

³⁵Esto solo es válido tomando en cuenta la segregación de los pobres, pues si existe evidencia de hiper-segregación (Duncan mayor a 0,6) del sector afluente (Valenzuela et al., 2008).

índice, imitando los trabajos de Valenzuela et al. (2013) y Paredes et al. (2013) respectivamente, e incorporando el cálculo del *rank-order segregation index*. Un análisis de la segregación a nivel de comuna, tanto escolar como urbana, se desarrolla en el anexo C.

3. DATOS

La base de datos utilizada proviene de diversas fuentes. Las dos principales son el SIMCE II° Medio 2006 y PSU 2009, que permitieron obtener indicadores de desempeño escolar, características y preferencias de las familia, percepciones del establecimiento escolar, y las georeferencias de los alumnos.

3.1 Fuentes Principales

La base SIMCE 2006 para 2° Medio incluye detalles de resultado de pruebas estandarizadas¹, y datos socioeconómicos de la familia, información sobre sus preferencias, y percepciones acerca del establecimiento educacional en su módulo Cuestionario para Padres y Apoderados. Este último lo responden los apoderados en sus hogares y no incluye la dirección de los alumnos.

La base PSU corresponde a los resultados y cuestionarios relacionados a la “Prueba de Selección Universitaria”, el instrumento oficial para el proceso de admisión a la educación superior. Se puede realizar una vez terminada la enseñanza media, y es de carácter voluntario, aunque la gran mayoría de los egresados de cuarto medio la inscriben. Ésta incluye la dirección del lugar de residencia y, asumiendo que los alumnos no se han cambiado de domicilio entre el 2006 y el 2009, permite tener una referencia de su ubicación en el 2006²

¹Prueba realizada a nivel nacional en todos los establecimientos que imparten educación media reconocidos por el Ministerio de Educación. Acorde al Manual de Uso de la Base de Datos SIMCE 2006, la prueba “... evaluó los Objetivos Fundamentales y Contenidos Mínimos Obligatorios (OF-CMO) del Marco Curricular correspondientes al Primer y Segundo año de la Enseñanza Media”.

²Al no existir la totalidad de los estudiantes en ambas bases, no es posible realizar un correcto pareamiento de todas las observaciones de la base SIMCE con la PSU, debido cuatro razones: (1) repitencia, (2) RUNs inválidos o duplicados en base SIMCE, (3) datos nulos y (4) no inscritos en PSU. A nivel nacional, de las 244.594 observaciones del SIMCE 2006 solo es posible vincular con datos PSU 2009 161.716 observaciones, más detalles e implicancias en Bravo et al. (2010). Además, aquellas observaciones que realizaron el SIMCE en un establecimiento de la Región Metropolitana, y que rindieron la PSU en una región distinta,

Se consideran las observaciones sin valores omitidos en variables relevantes para la estimación, que atiendan a colegios con financiamiento público o privado del Gran Santiago³. Suman un total de 37.859 observaciones distribuidas en 798 colegios.

3.2 Fuentes Auxiliares

Otras fuentes que se utilizaron:

CASEN: : Se utilizaron las base de datos de las encuestas de Caracterización Socioeconómica Nacional (CASEN) del Ministerio de Planificación (hoy Ministerio de Desarrollo Social) correspondientes a los años 2003, 2006 y 2009 para obtener las distribuciones de ingreso por colegio, y construir las restricciones de disponibilidad. Además, se utilizó la georeferencia de la CASEN 2003 (construida en Chumacero et al., 2011) para imputar datos de ingreso en base a ubicación dentro de la ciudad. Se corrigieron los datos de años distintos al 2006 de acuerdo a la variación del ingreso mensual promedio per cápita según Encuesta Suplementaria de Ingresos del INE.

Listado colegios católicos: : Lista que contiene los colegios que se declaran católicos al año 2008⁴.

Cobro de mensualidad colegios privados: : Base construida para este trabajo debido a que el dato está truncado en la declaración de pago de mensualidad del Cuestionario para Padres y Apoderados⁵. De los 177 colegios privados que se

no sirven para el propósito del estudio, pues su dirección declarada es obviamente distinta a la que tenían en segundo medio.

³Ver anexo A para mayores detalles de conformación de la base.

⁴Agradecimientos al Hno. Aldo Passalacqua por compartir la base.

⁵La categoría más alta para la pregunta “¿Cuánto dinero paga como mensualidad en el establecimiento?” corresponde a “Más de \$100.000”, lo que implica que solo puede estar truncado el pago en caso de colegio privado, pues el cobro máximo de un colegio con subvención estatal es de aproximadamente 4 USE mensual: \$54.020 para el 2006.

consideran en la estimación, 152 datos de cobro se obtuvieron a través de contacto telefónico directo con el colegio⁶, 20 fueron datos entregados por actuales alumnos o ex-alumnos, y para cinco casos no se encontró información. Se corrigieron los datos según la variación del IPC entre marzo 2006 y marzo del año correspondiente al dato.

Matrícula: : Base oficial de matrícula a nivel de colegio año 2006 elaborada por el Ministerio de Educación⁷.

Rendimiento: : Base oficial de rendimiento (notas) a nivel de alumno para los años 2002 a 2006, recopilada por el Ministerio de Educación⁸.

3.3 Georeferenciación y Distancia

La georeferenciación de los hogares fue realizada para el trabajo de Chumacero, Gallagos, y Paredes (2014). Para el caso de los colegios se utilizó la misma fuente, con la excepción de verificar las georeferencias publicadas por el Ministerio de Educación en sus bases de matrícula⁹. Se calculó la distancia euclidiana entre los colegios y los individuos, como *proxy* de la distancia real.

De los cálculos expuestos en la tabla 3.1, se puede apreciar que la distancia promedio al colegio es de 4 km, el 9 % de los alumnos asiste al establecimiento más cercano de su casa, mientras que el 25 % a uno de los cinco más cercanos. Respecto a los estudiantes de establecimientos municipales, los de colegios privados viajan, en promedio, distancias un 35 % menores y los de establecimientos subvencionados un 40 % menores. Un resultado interesante es que casi dos tercios de los alumnos de colegios municipales no viven en la comuna del establecimiento, mientras que más de la mitad de los matriculados en colegios subvencionados no viajan fuera del municipio.

⁶No se pudo obtener la información de todos los colegios por este medio debido a: (1) no entregan información por medio telefónico, (2) no se pudo establecer contacto, (3) el colegio ya no existe.

⁷Disponible en <http://centroestudios.mineduc.cl>.

⁸Disponible en <http://centroestudios.mineduc.cl>.

⁹Utilizando las georeferencias del Ministerio de Educación y *Google Maps* se corrigieron las ubicaciones de 32 colegios.

Un análisis detallado a nivel de comuna muestra que hay municipios como Santiago, Providencia o La Cisterna que atraen muchos alumnos; otros en los que la gran mayoría emigra (como Lo Espejo, Lo Prado y Cerro Navia); y algunos con condiciones mixtas en que llegan bastantes estudiantes de otras comunas, pero los alumnos locales viajan fuera a estudiar, como San Joaquín o Quinta Normal. Para un análisis completo referirse al anexo C.

TABLA 3.1. Cálculos en base a las distancias a los colegios

	Dependencia			Todos
	Municipal	Subvencionado	Privado	
Distancia Promedio [km]	5,5	3,3	3,7	4,0
Asiste al más cercano [%]	9,0	9,3	9,3	9,2
Asiste al más cercano de su dependencia [%]	13,7	13,5	12,7	13,4
Asiste a uno de los 5 más cercanos [%]	24,5	24,3	24,9	24,5
Asiste a uno de los 15 más cercanos [%]	41,6	41,2	42,0	41,4
Asiste dentro de su municipalidad [%]	35,0	56,9	48,5	49,4

3.4 Variables a Nivel Individual

Las variables a nivel individual se construyeron a partir de las declaraciones de los apoderados en el Cuestionario para Padres y Apoderados del SIMCE II° Medio 2006¹⁰. Se definió las siguientes variables:

Ingreso per cápita:: La declaración de ingreso familiar es por rangos¹¹, por lo que se definió la media del intervalo como el ingreso respectivo. Se dividió por el número de personas en el hogar. Para los ingresos declarados en el último tramo se imputó el ingreso como un promedio simple entre el ingreso de sus vecinos

¹⁰Para completar algunos valores en blanco en educación de los padres se utilizó el módulo de preguntas a los padres de la base PSU 2009.

¹¹Selección múltiple entre 13 alternativas, con rangos de ancho \$100.000 hasta los \$600.000 y de ancho \$200.000 hasta \$1.800.000, siendo la menor “Bajo \$100.000” y la más alta “Sobre \$1.800.000” (para los que se definió \$50.000 y \$1.900.000 de ingreso a los pertenecientes a esos rangos respectivamente).

(acorde a Casen 2003) y el promedio del ingreso de familias del mismo colegio con padres de igual nivel educacional¹².

Años educación madre:: Cantidad de años de educación alcanzados por la madre¹³.

Años educación padre:: Cantidad de años de educación alcanzados por la padre¹⁴.

Mujer:: Variable *dummy* que identifica las observaciones de sexo femenino.

Expectativas:: Variables *dummies* excluyentes que indican si el apoderado tiene expectativas que el pupilo termine educación superior técnica o educación universitaria.

Preferencias:: Variables *dummies* que indican si una de las tres principales razones de la elección del colegio es por la entrega de valores, por el prestigio del establecimiento, por la cercanía, por el logro en SIMCE, por el logro en PSU, por la adecuación al presupuesto familiar, por el nivel socioeconómico y cultural del alumnado, o porque miembros de su familia estudian o estudiaron en el establecimiento.

La tabla 3.2 muestra las medias según decil de ingreso per cápita. Se observa que existen al menos un 10 % más de observaciones de sexo femenino para los tres primeros deciles¹⁵, diferencia que se reduce para grupos con mayor ingreso. Entre un 21 % y 24 %

¹²Esto se realizó para 2911 observaciones: 7,7 % de la muestra, ver anexo B para mayores detalles.

¹³Se utilizó la conversión utilizada por el Ministerio de Educación en el *Manual de Uso de la Base de Datos SIMCE 2006* (basado en escala creada por la UNESCO). Para 109 casos (0,3 % de la muestra) que no se tiene información, se imputó los años de educación de la madre con los años de educación del padre.

¹⁴Para 1136 casos (3 % de la muestra) que no se tiene información, se imputó los años de educación del padre con los años de educación de la madre.

¹⁵Una razón puede ser que las tasas de egreso en grupos vulnerables son favorables a las mujeres (Raczynski et al., 2011), lo que explica mayor proporción de sexo femenino pues todas las observaciones de la base inscribieron la PSU, por ende estuvieron en cuarto medio.

de padres de los seis primeros deciles creen que sus hijos van a completar educación superior en un CFT o IP¹⁶, bajando a 15 % en el séptimo decil y a 5 % o menos para los dos últimos. Con una tendencia contraria, el promedio de apoderados con expectativas universitarias crece monótonicamente de 37 % a 98 % acorde al ingreso. Padres con mayores recursos tienden a priorizar en la elección de colegio la formación de valores, el prestigio del establecimiento y el nivel socioeconómico del alumnado con mayor frecuencia que grupos más pobres; y a poner menor atención en el SIMCE, costo y distancia (esto último es relativamente parejo para nueve décimos de la población, decayendo en importancia únicamente para el decil más afluente).

En promedio, los apoderados no declaran que el SIMCE ni la PSU están entre las tres razones más importantes para la elección de colegio (solo lo hace un 7 % y un 9 % respectivamente); siendo las más relevante la cercanía y el prestigio, consideradas por un 43 % de las observaciones. La percepción de los padres¹⁷ sobre la calidad, la gestión y la información entregada por el colegio sobre el SIMCE, es bastante pareja para los nueve primeros deciles, y sube sutilmente para el último. La percepción promedio sobre el desempeño docente es levemente decreciente en ingreso hasta el decil más rico, que experimenta un alza hasta casi igualar la percepción de los más pobres. Se observa que a mayor ingreso hay bastante mayor conocimiento de resultados SIMCE de años anteriores.

3.5 Variables a Nivel Colegio

La mayoría de las características de los colegios se definieron a partir de las declaraciones de los apoderados en el Cuestionario para Padres y Apoderados del SIMCE IIº Medio 2006 pertenecientes a cada establecimiento:

Ingreso per cápita promedio:: Media del ingreso per cápita de los alumnos de la base.

¹⁶CFT: Centro de Formación Técnica, IP: Instituto Profesional.

¹⁷Referida al colegio que asiste su hijo.

TABLA 3.2. Medias de variables nivel individual por decil de ingreso

Decil	1	2	3	4	5	6	7	8	9	10	Todos
Distancia [km]*	2.3	2.5	2.5	2.7	2.8	2.8	2.7	2.8	2.6	2.8	2.6
Copago [2006 M\$]	9	10	12	13	16	18	26	42	96	197	43
Ingreso per cápita [2006 M\$]	18	31	39	51	65	85	111	158	281	714	153
Educación madre [años]	9.6	10.4	10.7	11.2	11.8	12.2	13.1	14.1	15.3	16.4	12.4
Educación padre [años]	9.8	10.6	10.7	11.4	11.9	12.4	13.3	14.5	16.0	17.1	12.7
Mujer [%]	59	55	55	57	53	53	52	53	52	51	54
Tiene auto [%]	15	22	22	33	41	46	63	77	90	98	50
Expectativas educacionales											
Carrera técnica [%]	22	24	24	24	23	21	15	9	5	2	17
Carrera universitaria [%]	37	45	49	56	64	69	80	88	94	98	67
Preferencias por											
Valores [%]	26	28	29	29	30	31	34	36	48	66	36
Prestigio [%]	30	36	38	39	42	45	45	49	47	56	43
Cercanía [%]	46	45	45	43	42	42	43	42	45	36	43
Nivel socioeconómico [%]	3	4	4	5	6	6	8	10	13	19	8
SIMCE [%]	6	6	6	7	7	9	9	10	8	4	7
PSU [%]	4	5	6	7	8	10	12	14	12	11	9
Costo [%]	40	43	40	42	44	41	42	38	32	17	38
Tradición [%]	29	27	27	26	24	25	23	24	26	35	27
Percepción por atributos del colegio [escala de 1 a 4]											
Calidad	2.99	2.99	3.00	2.97	2.97	2.97	2.97	2.96	3.01	3.18	3.00
Información sobre SIMCE	2.71	2.74	2.74	2.70	2.74	2.72	2.73	2.72	2.73	2.82	2.73
Gestión	2.74	2.76	2.76	2.74	2.77	2.75	2.74	2.74	2.81	3.01	2.78
Profesores	3.07	3.05	3.04	3.02	3.00	2.99	2.97	2.95	2.96	3.05	3.01

*:mediana.

Desviación estándar ingreso per cápita:: Desviación estándar del ingreso per cápita de los alumnos de la base .

Educación Madres:: Media de los años de educación de la madres de los alumnos de la base.

Copago:: Mediana del pago de mensualidad declarado. Para los colegios privados, se corrigió el valor según base propia (base de cobro de mensualidad colegios privados, ver sección 3.2).

Proceso Admisión:: Variables *dummies* que representan si al menos $\frac{2}{3}$ de las observaciones del colegio declaran que le exigieron al matricularse: certificado de

bautizo y/o matrimonio por la Iglesia, entrevista con los padres, examen escrito o prueba de ingreso, o certificado de remuneraciones.

Satisfacción:: Promedio del nivel de satisfacción declarado (escala de 1 a 4) sobre la calidad de educación, infraestructura e instalaciones, desempeño de los profesores, y formación valórica y moral.

Ingreso vecindario:: ¹⁸ Ingreso promedio¹⁹ de los vecinos del colegio.

Y de otras fuentes se definieron:

SIMCE:: ²⁰ De la base de resultados a nivel alumno del SIMCE II° Medio 2006 y 2004 se calculó la media entre las pruebas de Lengua Castellana y Matemática de ambos años.

Católico:: Del Listado de Colegios Católicos se construyó una variable *dummy* que representa si el colegio se declara católico.

Nombre en inglés:: Se definió una variable *dummy* para los colegios cuyo nombre contuviera alguna palabra en Inglés²¹.

Tamaño:: De la base de matrícula 2006 se obtuvo la cantidad de alumnos matriculados en educación media²².

Tipo de enseñanza:: Variables *dummies* si el colegio ofrece enseñanza de tipo científico-humanista (C-H) y/o técnico profesional (T-P).

La tabla 3.3 describe los colegios según dependencia, acorde a las variables recién presentadas. Se observa que, en promedio, los colegios subvencionados tienen SIMCE

¹⁸Esta variable intenta ser un *proxy* a nivel de vulnerabilidad o “seguridad del barrio” en que se ubica el colegio, factor importante para la elección de acuerdo a Raczynski et al. (2010).

¹⁹Promedio ponderado por $\frac{1}{(distancia+100)^3}$ de los vecinos en un radio 750m.

²⁰Es el *proxy* más importante para calidad del establecimiento que se incluye. Es discutible su capacidad para representarla, sobre todo por su descomposición no observable entre el aporte real del colegio y la “calidad” del alumnado; algo probablemente difícil de identificar por los electores (Grau, 2014).

²¹Las palabras se limitaron a: *school, college, international, saint, academy, English, institute, British*.

²²Si bien esta variable puede reflejar una “profecía autocumplida”, Raczynski et al. (2010) destacan la importancia que le dan los apoderados a la cantidad de demanda que tienen los colegios, asociándola con un símbolo de calidad.

más alto que los municipales y más bajo que los particulares; esta última relación también se da en el copago medio y en el nivel de educación e ingreso de los apoderados. Un 33 % de los colegios municipales exige prueba o examen de ingreso, subiendo a 46 % para subvencionados y 78 % para privados. Más grande es la diferencia en los colegios que exigen entrevista con los padres: 13 % de los municipales, 31 % y 89 % de los subvencionados y particulares respectivamente ²³. La percepción promedio sobre formación valórica es la que presenta mayor diferencia entre dependencias, y la de desempeño de los profesores la que menos, variando de 2,98 a 3,04 (en escala de 1 a 4).

Un tercio de los colegios municipales ofrece solo enseñanza técnico-profesional, mientras que un 43 % solo dicta currículum científico-humanista. En el sector subvencionado solo un 18 % entrega exclusivamente educación técnico-profesional, mientras que un 13 % es polivalente. En la educación privada no se imparte currículum técnico-profesional.

²³Las prácticas anteriores no necesariamente están relacionadas con la decisión de admitir al postulante. Por ejemplo, el caso de exigir examen no significa que se selecciona condicional al resultado, está la posibilidad que se utilice el instrumento como un diagnóstico. Más discusión en Carrasco et al. (2014), un estudio del proceso de admisión y la selección en el sistema escolar chileno, a partir de declaraciones de los directores.

TABLA 3.3. Medias de variables nivel colegio por dependencia

Dependencia	Municipal		Subvencionado		Particular		Todos	
SIMCE	239	(33)	255	(32)	304	(30)	263	(39)
Copago [2006 M\$]	2	(5)	23	(16)	175	(60)	53	(73)
Media educación madres [años]	10.5	(1.7)	12.0	(2.0)	15.9	(0.8)	12.6	(2.5)
Desv. Est. Ingreso [2006 M\$]	39	(23)	62	(36)	210	(119)	91	(90)
Media Ingreso [2006 M\$]	56	(27)	93	(52)	444	(205)	164	(184)
Nombre en Inglés [%]	0		9		21		10	
Católico [%]	0		28		38		25	
Tamaño	755	(492)	429	(446)	222	(167)	446	(446)
Ingreso vecindario [log(\$)]	11.32	(0.98)	11.20	(0.90)	12.26	(1.06)	11.46	(1.05)
Requisitos de admisión								
Examen [%]	33		46		78		50	
Entrevista [%]	13		31		89		40	
Cert. Religioso [%]	0		12		26		13	
Notas [%]	83		77		38		69	
Percepciones sobre [escala de 1 a 4]								
Valores	2.82	(0.30)	3.02	(0.35)	3.27	(0.39)	3.03	(0.38)
Infraestructura	2.66	(0.44)	2.75	(0.53)	2.91	(0.54)	2.77	(0.52)
Calidad	2.82	(0.31)	2.90	(0.31)	3.06	(0.32)	2.92	(0.32)
Profesores	2.98	(0.25)	2.99	(0.27)	3.04	(0.27)	3.00	(0.27)
Tipo de enseñanza								
Técnico-profesional [%]	57		31		0		29	
Científico-humanista [%]	67		82		100		83	

Desviación estándar entre paréntesis.

4. CÁLCULOS DE SEGREGACIÓN

Utilizando la base de datos descrita en el capítulo anterior, se calculó la segregación con diferentes índices y criterios. Se consideró el índice Duncan, por su popularidad en la investigación nacional, y el índice de teoría de la información, por sus propiedades y sustento teórico.

En la Figura 4.1 se muestra una representación gráfica de la segregación según cotas de vulnerabilidad definidas por percentil (de 10 a 90) de la distribución de ingreso para estos dos índices. Para ambos gráficos, las abscisas corresponden al límite de “vulnerabilidad” que se utiliza para calcular el índice, y las ordenadas a la magnitud de segregación calculada. Por ejemplo, el valor que representan las curvas cuando x es igual a 0,2, es la segregación si se considera como “vulnerables” a todas las observaciones con ingreso menor o igual al percentil 20 de la distribución de ingreso per cápita¹.

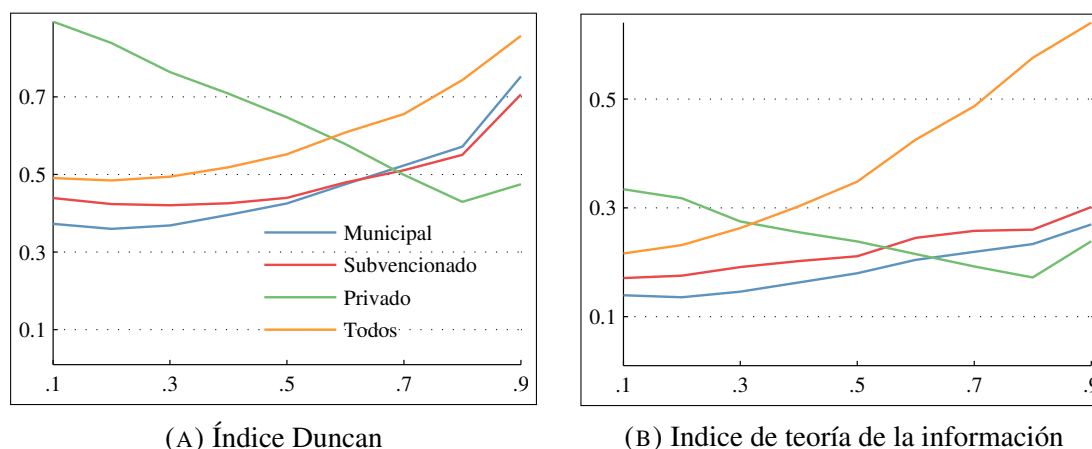


FIGURA 4.1. Índices de segregación (eje Y) para distintos percentiles de vulnerabilidad (eje X).

Se observa que los perfiles de segregación son similares, a excepción de la segregación total, que en el caso de índice H crece con mayor pendiente a medida que se considera

¹Que es equivalente a considerar la segregación del 80 % más afluyente, pues todos los índices calculados en este trabajo son simétricos.

un grupo “vulnerable” con mayor ingreso. La diferencia se podría explicar en parte por la manera en que los índices “castigan” el aporte marginal a la segregación, detallado en la sección 2.2.1. La segregación es mayor mientras se considere un grupo “vulnerable” con más recursos, lo que es válido para todos menos el sistema privado, independiente del índice que se evalúe.

Considerando el índice Duncan, el decil más rico está un 70 % más segregado que el decil más pobre, mientras que según el índice de teoría de la información está un 200 % más segregado. El sistema municipal es el más integrado, al menos considerando grupos hasta el sexto decil. Los más ricos están mejor integrados en el sector privado, pero considerando H la diferencia es leve respecto a su integración en municipales y subvencionados.

4.1 Descomposición *between* y *within*

La descomposición de un índice ayuda a entender a qué nivel se produce la segregación, pues permite observar si la segregación ya se manifiesta analizando grupos de colegios (separados por ejemplo por tipo de colegio o nivel de cobro), o si es la distribución de los estudiantes entre los establecimientos dentro de los grupos lo que explica gran parte de la segregación.

Para poder realizar la descomposición es necesario agrupar colegios. Por ejemplo, si se clasifica por dependencia, los 798 colegios se dividirían en tres grupos: municipales, subvencionados y particulares, con lo que es posible calcular la segregación *between* y *within*. La primera es la segregación que existe considerando los grupos como unidad análisis, lo que equivale a que cada grupo se convierta en un gran colegio. Si cada grupo tiene exactamente la misma proporción de “vulnerables”, independiente de cómo se distribuyan los estudiantes en los colegios reales dentro del grupo, la segregación *between* será cero.

Por otro lado, la segregación *within* es un promedio ponderado² de la segregación calculada de forma aislada para cada grupo, utilizando los colegios como unidad de análisis, y dada la proporción de “vulnerables” del grupo (independiente de si esta proporción es mayor o menor que la de otros grupos). En otras palabras, se considera cada grupo como un sistema aparte y se calcula qué tan segregados están los colegios dentro del grupo. La segregación será cero si los colegios presentan la misma proporción de “vulnerables” que el grupo tomado en su conjunto.

En resumen, la segregación *between* se refiere a cómo se distribuyen los “vulnerables” entre los grupos, y la *within* al interior de cada uno de los grupos (condicional a la repartición entre grupos).

4.1.1 Índice para variables dicotómicas

El procedimiento de descomposición se puede realizar para todos los índices reportados en la sección 2.2, con excepción del índice Duncan (Hutchens, 2004; Reardon y Firebaugh, 2002). En la Tabla 4.1, se muestran los resultados de diferentes divisiones propuestas, utilizando el índice de teoría de la información, y considerando el grupo vulnerable como el 30 % de menor ingreso per cápita. Se dividió por dependencia, religión, nivel de cobro y tipo de currículum impartido.

Se aprecia que la segregación *between* es bastante importante en la descomposición por copago. Si es que la integración fuera perfecta en cada uno de los grupos de colegios de igual cobro (sin cambiar a nadie del grupo, solo de colegio dentro de su mismo grupo), la segregación todavía sería un 58 % de la original. Se observa que el sistema subvencionado

²Ponderado por la cantidad de observaciones y el nivel de diversidad de cada colegio (Reardon y Firebaugh, 2002). Lo primero es coherente con que el índice está basado en las personas y no en los territorios (en cuanto a importancia), y lo segundo tiene sentido si se toma en cuenta que es más probable que un grupo con una proporción de “vulnerables” baja (o sea un grupo poco diverso) esté muy segregado, y esto no debería pesar en gran medida en la segregación final (esto sí influye en la segregación *between*, lo relevante en ese caso es que haya baja proporción de minoritarios, no que estén segregados, pues si son poco es altamente probable que lo estén).

TABLA 4.1. Descomposición del índice de teoría de la información (H).

	Dependencia			Religión		Copago				Currículum		
	Municipal	Subvencionado	Particular	Católico	Otros	≤ 7.500	≤ 17.500	≤ 60.000	> 60.000	Solo T-P	T-P y C-H	Solo C-H
Colegios [%]	20	58	22	25	75	32	19	27	22	17	12	71
Observaciones [%]	28	56	16	26	74	38	20	26	16	21	16	63
“Vulnerables” [%]	43	32	1	20	34	46	41	17	1	53	50	17
Segregación*	146	191	275	295	239	139	96	147	275	77	94	253
Aporte a seg. <i>within</i> [%]	29	69	2	25	75	53	19	25	2	12	11	77
Segregación <i>within</i> *	158 (60 %)			247 (94 %)		111 (42 %)				155 (59 %)		
Segregación <i>between</i> *	105 (40 %)			16 (6 %)		151 (58 %)				108 (41 %)		
Segregación total*	263											

*Índices de segregación multiplicados por 1000.

T-P: técnico profesional, C-H: científico humanista.

es el que más aporta a la segregación *within*, aunque en el particular están más segregados los pocos “vulnerables” que lo componen.

Existe mayor segregación entre los colegios católicos que los no católicos, pero el aporte a la segregación total es, en proporción al tamaño, parecida para ambos grupos³. Un resultado interesante es que los colegios con copago entre \$7.500 y \$17.500 están menos segregados que los establecimientos con mensualidades menores a ese rango⁴. También se observa que la segregación en los colegios que imparten currículum técnico profesional es bastante menor que la que solo ofrecen educación científico-humanista.

³Para un estudio específico de la selección en colegios católicos ver Madero y Madero (2012).

⁴Una posible explicación sería que el público de colegios con menor cobro es el más restringido a viajar, y probablemente el más segregado en términos residenciales, lo que se transmite de forma más directa en la segregación escolar. Otra es que este grupo puede ser más diverso en términos de oferta educativa y perfil de estudiantes, pues dentro de estos colegios están los liceos de excelencia, lo que hace más probable la segregación. Son solo ideas que requieren mayor estudio.

El análisis anterior se refiere únicamente a la segregación del 30 % con menor ingreso. Una mirada más global se obtiene a partir de los gráficos 4.2A y 4.2B. Éstos consideran la segregación (ordenadas) según grupos “vulnerables” definidos con distintos percentiles de ingreso (abscisas). La curva amarilla representa la segregación total, el área pintada la segregación *within*, con el aporte específico de cada subgrupo por color, y la diferencia o espacio en blanco, la segregación *between*.

De la Figura 4.2A se observa que el aporte a la segregación de los colegios subvencionados es el más importante para todos los límites de vulnerabilidad, excepto del 10 % más afluente. Si bien es el grupo más grande de colegios, la contribución sigue siendo la mayor incluso proporcional al tamaño. De la Figura 4.2B se concluye que el grupo de colegios con copago menor a \$7.500 es el que más influye en la segregación *within*, salvo cuando se analiza el grupo con más recursos.

Uno de los resultado más sorprendente de ambos gráficos es que la falta de integración se explica en mayor parte por la segregación *between*, es decir por la elección de un colegio de un grupo: municipal, subvencionado o privado, o de cierto rango de mensualidad, y en menor parte por la elección del colegio específico dentro de esos grupos⁵. Más fuerte es el fenómeno para el caso de la división por cobro, y puede que se transmita a las dependencias por correlación⁶. El efecto *between* es cada vez más relevante si se considera la segregación de un grupo de mayor ingreso.

Por último, para evaluar si la gran importancia del efecto *between* es característica del indicador, se calculó utilizando otros grupos. En la Figura 4.3 se muestran los resultados para divisiones según religión, requisitos de admisión y tamaño; donde claramente el efecto *within* domina al *between*, descartando que el indicador “tienda” a sugerir mayor

⁵Lo que no quiere decir que la elección sea necesariamente secuencial, sino que existe correlación entre la segregación y estas divisiones.

⁶Raczynski et al. (2010) sugieren que las familias más que diferenciar colegios por tipo de dependencia, lo hacen por nivel de cobro.

segregación entre grupos que de dentro de los grupos⁷. Esto indica que las principales razones de la segregación debieran estar correlacionadas fuertemente con el nivel de copago y, también pero en menor medida, con la dependencia.

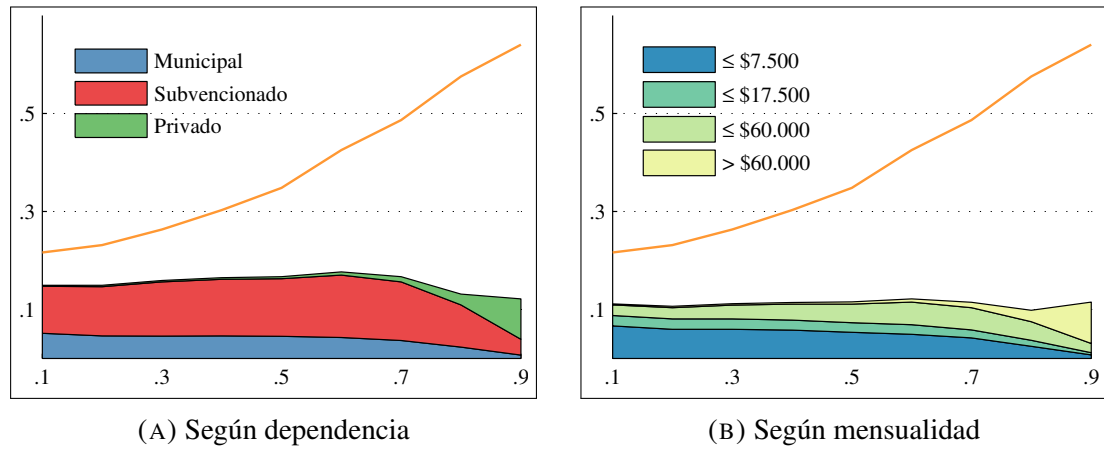


FIGURA 4.2. Descomposición de H (eje Y) para distintos percentiles de vulnerabilidad (eje X)

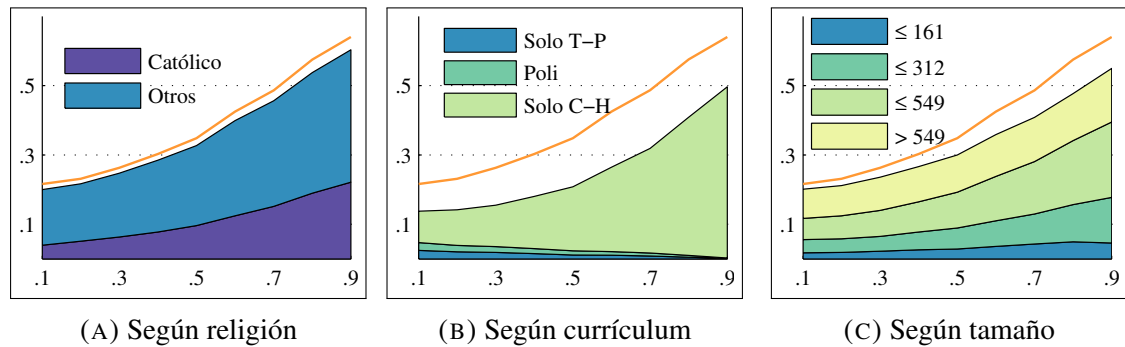


FIGURA 4.3. Otras descomposiciones de H (eje Y) para distintos percentiles de vulnerabilidad (eje X)

⁷Se construyeron estos gráficos utilizando los índices de tasa de varianza y de raíz cuadrada, llegando a los mismos resultados (interesados, solicitar al autor).

4.1.2 Índice para variables continuas

Como novedad, se calcula la descomposición de índices de segregación que no consideran a un grupo como “vulnerable”, sino que miden la integración considerando una variable continua, como lo es el ingreso⁸. Utilizando los mismos grupos de la sección anterior, los resultados se muestran en la Tabla 4.2.

TABLA 4.2. Descomposición del *rank-order information theory index (R-O H)*

	Dependencia			Religión		Copago				Currículum		
	Municipal	Subvencionado	Particular	Católico	Otros	≤ 7.500	≤ 17.500	≤ 60.000	≥ 60.000	Solo T-P	T-P y C-H	Solo C-H
Colegios [%]	20	58	22	25	75	32	19	27	22	17	12	71
Observaciones [%]	28	56	16	26	74	38	20	26	16	21	16	63
Segregación*	186	222	246	400	359	178	141	163	246	116	141	351
Aporte a seg. <i>within</i> [%]	25	66	9	31	69	42	16	30	12	5	5	90
Segregación <i>within</i> *	154 (40 %)			358 (94 %)		112 (29 %)				250 (65 %)		
Segregación <i>between</i> *	228 (60 %)			24 (6 %)		270 (71 %)				133 (35 %)		
Segregación total*	382											

*Índices de segregación multiplicados por 1000.

T-P: técnico profesional, C-H: científico humanista.

Los resultados imitan lo que sería un promedio de tablas como la 4.1, considerando distintos niveles de “vulnerabilidad”, que sean representativos de la toda la distribución (no solo el percentil 30). También equivale a “promediar horizontalmente” los resultados de la Figura 4.2⁹.

Los resultados no difieren en gran medida de los de la Tabla 4.1, descritos en la sección 4.1.1, que representa la segregación considerando el percentil 30 de la distribución de

⁸Más detalles de este índice se desarrollan en sección 2.2.2. Construcción y procedimiento de descomposición se describen en anexo D.

⁹En la práctica sería integrar. Ambos promedios ponderados por el nivel de diversidad del sistema.

ingreso per cápita como límite de “vulnerabilidad”. Los cambios más relevantes son el aumento de importancia de la segregación *between* sobre la segregación total. Considerando este nuevo índice, la segregación *between* explica el 59 % de la segregación si se divide por dependencia, y un 71 % si se separa por nivel de cobro. Nuevamente, esto quiere decir que si dentro de cada grupo los individuos pertenecientes estuvieran perfectamente integrados¹⁰, la segregación del sistema se mantendría en un 59 % y un 71 % de la original para las dos divisiones recién mencionadas.

También aumenta la incidencia de los colegios católicos en la segregación, bajo este índice es mayor el aporte por colegio de los católicos que el aporte de los no católicos¹¹. La configuración del aporte a la segregación *within* en la división por tipo de currículum cambia levemente, los colegios con únicamente opción científico-humanista tienen mayor incidencia en la segregación escolar.

La segregación del sistema, considerando la distribución individuos de distinto ingreso en los colegios (*rank-order information theory index*), es un 45 % más alta en comparación a la del grupo dicotómico constituido por el 30 % de menor ingreso (índice de teoría de la información). La explicación es simple, grupos de mayor ingreso están más segregados que los de menor, independiente del índice que se utilice, lo que incide en el mayor valor del *R-O H* comparado con *H*. Si los individuos prefieren estar con pares similares o de mejor nivel socioeconómico (Raczynski et al., 2010), entonces los resultados anteriores demuestran la gran capacidad del sistema educacional chileno para lograr este objetivo mientras más recursos posea la familia.

En un contexto de “libre” elección, esto implica que la perjudicial segregación de los más pobres (Larrañaga y Sanhueza, 2007) se deba a que los individuos de mayor ingreso,

¹⁰Como ya se ha descrito, para el concepto de segregación que describe el *rank-order information theory index*, esto implica que cada unidad (dentro del grupo) tenga una distribución similar de individuos considerando nivel de ingreso per-cápita. Y si es así, entonces por construcción tendrían la misma distribución del grupo.

¹¹Este resultado se desprende de dividir el aporte a la segregación *within* por el porcentaje de las observaciones pertenecientes a cada grupo.

en su derecho de optar por un colegio de acuerdo a sus preferencias, en promedio acuden a establecimientos a los que familias de menos recursos tienen restricción de acceso¹². Este fenómeno incrementa con el ingreso, mientras se considere familias más afluentes, menor es la posibilidad de que compartan la sala de clases con individuos de ingreso inmediatamente menor.

Los datos recién expuestos no permiten identificar bien los mecanismos o razones que hacen un colegio accesible para algunos y restringido para otros, y que en definitiva producen la segregación. Pero advierten que un sistema educacional integrado no se logra solucionando únicamente la segregación de los vulnerables, el problema es cada vez más grave en grupos con más recursos.

¹²O una hipótesis alternativa: que las preferencias por colegio sean cada vez más distintas al resto a medida que aumenta el ingreso familiar.

5. MODELACIÓN Y ESTIMACIÓN

La información que entregan los índices de segregación es importante para el diagnóstico general, pero es limitada la ayuda si se busca entender las razones o mecanismos que producen la segregación. En este trabajo se aborda el problema desde el proceso de elección de colegio. En el sistema educacional chileno las familias son “libres” de elegir el establecimiento escolar¹, lo que implica que la segregación se puede explicar en parte por la compleja interacción de miles de elecciones de establecimiento escolar. Este capítulo describe la estructura elegida para estimar las preferencias de las familias que definen las decisiones de elección de un colegio, y resume brevemente el novedoso proceso de estimación utilizado.

5.1 Modelación

El proceso de elección de un establecimiento educacional es discreto, es decir, no se elige una cantidad parcial de una alternativa y un poco de otra, si no una única opción entre todas las disponibles. Para estudiar estos escenarios se recurre a modelos de variable dependiente discreta, siendo el logit multinomial (McFadden, 1974) el más popular, y la base de numerosas extensiones. En forma general, estos modelos permiten obtener una probabilidad de elección de las distintas opciones disponibles, lo que se logra ajustando un conjunto de parámetros que representan la valoración individual por los atributos de los colegios, de manera de maximizar la probabilidad predicha para las elección real².

¹En la medida que tengan la capacidad para cubrir los costos de matrícula y mensualidad, que estén de acuerdo con el proyecto educativo, y que además sean aceptados por el establecimiento (no todos los colegios pueden legalmente seleccionar, ver sección 1.1).

²Representar la elección de colegio con un modelo de elección discreta puede ser una especificación incorrecta, por lo que si no se tiene el cuidado necesario los parámetros resultarán sesgados. En la realidad, el proceso es mucho más complejo, pues las alternativas tienen un número limitado de cupos, y existen procesos de selección de diferente índole, por lo que el colegio preferido no es necesariamente igual al finalmente escogido. En el capítulo 6 se describe la estrategia utilizada que intenta minimizar el potencial sesgo en los coeficiente producto de la simplificación del proceso de elección.

El modelo propuesto sigue muy de cerca el trabajo de Gallego y Hernando (2008b), innovando en la incorporación de heterogeneidad en las preferencias de las familias por atributos de los colegios, sin imponer una forma funcional explícita que las relacione con las características familiares.

Esto se logra a través de un modelo logit de clases latentes (*latent class logit* o *finite mixture model*, Gupta y Chintagunta, 1994; Kamakura y Russell, 1989), el cual asume que las familias tienen preferencias distintas, asignándolas endógenamente dependiendo de sus características socioeconómicas a diferentes clases o segmentos, que representan diferentes patrones de preferencias³.

La asignación endógena de los individuos a los distintos segmentos está basada en que no se pueden identificar inmediatamente las clases a partir de los datos (Walker y Ben-Akiva, 2002), y pre-clasificar las observaciones, lo que genera la necesidad de estimar en conjunto los parámetros de pertenencia (probabilística) a cada clase, con los parámetros de preferencias por colegios. No se conocen más aplicaciones de este tipo a un modelo de elección de colegio.

³Esta nueva especificación no permite, a priori, realizar la estimación en dos etapas de Gallego y Hernando (2008b). Esto se debe a que no incluye constante específica por colegio; primero, porque la cantidad de parámetros crece en forma multiplicativa según el número de clases; y segundo, porque es altamente probable que generen problemas de identificación, dado que las clases están ligadas a características socioeconómicas de las familias, las constantes pueden ser clasificadores casi perfectos para ciertos grupos. Por ejemplo, una clase que represente la población más pobre, probablemente no tendrá demanda por colegios privados, lo que el modelo traducirá en constantes específicas “muy negativas” ($-\infty$) para estos colegios dentro de la clase. Cabe destacar que estas constantes específicas no aportan en la comparación de los criterios de elección según características de los demandantes, pues son iguales para todos. La probable consecuencia de su no inclusión es que el nivel de las preferencias por atributos de los colegios estimadas sin interacciones con atributos de los individuos, es decir las constantes (diferentes de las constantes específicas), estén sesgadas por correlación entre no observables y precio-calidad, tal como advierten Gallego y Hernando (2008b). Cabe mencionar que la diferencia en preferencias por características de los establecimientos entre distintos padres (cruces entre variables) debiera tener el mismo sesgo recién descrito, pues los parámetros de interacción de preferencia según características socioeconómicas no son estimados con variables instrumentales en el trabajo de los autores, ni en éste.

5.1.1 Teoría de utilidad aleatoria

El modelo logit de clases latentes se define bajo la teoría de utilidad aleatoria (McFadden, 1974). En su versión más general, supone que la utilidad U_{ij} de una alternativa j para un individuo i se deriva de un vector de atributos W_{ij} que describe cada alternativa, desde la perspectiva de cada individuo. Esta utilidad está compuesta por dos partes (ver 5.1). La primera es la determinística (V_{ij}), que contiene los atributos de la alternativa percibidos por el individuo (W_{ij}), y observables por el modelador, ponderados por la relevancia que cada individuo le asigna (λ_i). Y la segunda es la estocástica (ϵ_{ij}), que internaliza errores de medición y las características no observables por el modelador (Ortúzar y Willumsen, 2011).

$$\begin{aligned} U_{ij} &= V_{ij} + \epsilon_{ij} \\ &= W_{ij}\lambda_i + \epsilon_{ij} \end{aligned} \tag{5.1}$$

Si se define D_i como el conjunto de alternativas (colegios) disponibles para i , bajo esta teoría un individuo elige la alternativa j si le entrega mayor utilidad que cualquier otra que tiene disponible. Dado el componente aleatorio de la utilidad, se puede definir la probabilidad (P_{ij}) de que el individuo i elija el colegio j , tal como se muestra en (5.2).

$$\begin{aligned} P_{ij} &= P(U_{ij} \geq U_{ij'} \quad \forall j' \in D_i, \quad j' \neq j) \\ &= P(\epsilon_{ij'} \leq \epsilon_{ij} + V_{ij} - V_{ij'} \quad \forall j' \in D_i, \quad j' \neq j) \end{aligned} \tag{5.2}$$

Bajo el supuesto de que los errores ϵ_{ij} provienen de funciones idénticas e independientemente distribuidas (iid) *Gumbel*, y aprovechando que el máximo entre variables aleatorias iid *Gumbel* también distribuye *Gumbel*, se puede llegar a una expresión analítica de la probabilidad de la ecuación (5.2), que se exhibe en la ecuación (5.3): la definición del modelo logit (Walker y Ben-Akiva, 2002). Donde μ es el factor de escala que está relacionado con la desconocida varianza de los errores ϵ .

$$P_{ij} = \frac{\exp(\mu V_{ij})}{\sum_{\forall j' \in D_i} \exp(\mu V_{ij'})} \quad (5.3)$$

Los modelos logit están basados en maximización de utilidad a nivel individual (Bhat, 1997), por lo que sería ideal estimar un conjunto de parámetros (preferencias) λ_i para cada individuo. En la gran mayoría de los casos esto no es posible, pues no se tiene más que una observación por persona. Una alternativa es asumir homogeneidad en preferencias ($\lambda_i = \lambda \quad \forall i$), lo que puede ser un supuesto poco adecuado si se considera probable la existencia de variabilidad en gustos. Diversos trabajos en el contexto de elección discreta han resaltado la importancia de modelar la variabilidad, utilizando distintas aproximaciones para acomodar heterogeneidad en preferencias (Amador, González, y Ortúzar, 2005; Breffle y Morey, 2000; Hess y Polak, 2005; Provencher y Bishop, 2004).

Una de las formas más simples es agregar variación sistemática de gustos en un logit condicional (o MNL, McFadden, 1974), esto es, estimar parámetros que sean función de características de los individuos ($\lambda_i = f(Z_i)$, donde Z_i es un vector de característica de i , y f una función común a todos los individuos), como explican Ortúzar y Willumsen (2011). Una aproximación más compleja es a través de un modelo logit mixto (Train, 2009), el cual que los parámetros no son fijos, sino que siguen una distribución que representa la variabilidad en gustos ($\lambda \sim Dist(\bar{\lambda}, \sigma_\lambda)$). Hay algunas restricciones que podrían hacer de estos dos últimos modelos poco apropiados. El primero implica asumir una forma funcional $f(x)$ entre las preferencias y las características de los individuos (por ejemplo lineal en Gallego y Hernando, 2008b), que para la gran mayoría de los casos no es observable. Y el segundo requiere asumir cierta distribución ($Dist$) para los parámetros⁴.

⁴Otras restricciones de un logit mixto es que no agrega heterogeneidad directamente relacionada con variaciones sistemáticas en atributos de los individuos (Bhat, 1997). Si bien se pueden “recuperar” los parámetros individuales, no es trivial relacionar la heterogeneidad en preferencia con características de los individuos. Además, al imponer una distribución implica que existan realizaciones del parámetro difíciles de interpretar por magnitud o signo. Más discusión ver en Greene y Hensher (2003); Hess, Ben-Akiva, Gopinath, y Walker (2011); Provencher y Bishop (2004); Shen (2009).

5.1.2 Logit de clase latente

En este trabajo se utiliza una tercera alternativa, el modelo logit de clases latentes. Este modelo asume que existen S segmentos que representan las preferencias de diferentes tipos de individuos. En (5.4) se define la probabilidad de elección de la alternativa j para el individuo i condicional a su pertenencia a la clase s , donde λ_s son la preferencias específicas a la clase. Dentro de cada segmento los individuos tienen el mismo patrón de preferencias (no iguales preferencias, pues es posible agregar variación sistemática de gustos⁵) y el mismo nivel de certidumbre en la elección μ_s ⁶.

$$\begin{aligned} P_{ij|s} &= \frac{\exp(\mu_s V_{ij})}{\sum_{j' \in D_i} \exp(\mu_s V_{ij'})} \\ &= \frac{\exp(\mu_s W_{ij} \lambda_s)}{\sum_{j' \in D_i} \exp(\mu_s W_{ij'} \lambda_s)} \end{aligned} \quad (5.4)$$

Como no se sabe a qué clase pertenece cada observación, se utiliza una formulación logit condicional (5.5) para calcular la pertenencia a cada segmento, condicional a un vector M_i de características de los individuos que se decidan relevantes para la segmentación.

$$P_{is} = \frac{\exp(\mu M_i \theta_s)}{\sum_{s'=1}^S \exp(\mu M_i \theta_{s'})} \quad (5.5)$$

Con (5.4) y (5.5) se puede calcular (5.6): la probabilidad no condicional de la elección de la alternativa j por el individuo i , que no es más que la suma de las probabilidades de elegir la alternativa j en cada una de las clases, ponderada por la probabilidad de pertenecer a la clase. Definiendo A_j como el conjunto de individuos que elige realmente la alternativa

⁵Por lo que en estricto rigor, en Ecuación (5.4), $\lambda_s \Rightarrow \lambda_{is} = f_s(Z_i)$.

⁶Para la aplicación desarrollada en este trabajo se normalizan todos los μ_s a uno, pues para el logit de clases latentes en su versión tradicional los factores de escala no son identificables (Boxall y Adamowicz, 2002), si no más bien se “confunden” con los parámetros de preferencia. Existen extensiones que permiten levantar este supuesto, fuente de potencial sesgo (Greene y Hensher, 2010; Magidson y Vermunt, 2007), se deja para futuros desarrollos.

j , y J como el total de alternativas, se puede construir la función (5.7) de verosimilitud asociada al problema, definida como la probabilidad conjunta de las elecciones reales⁷.

$$P_{ij} = \sum_{s=1}^S P_{is} P_{ij|s} \quad (5.6)$$

$$\mathcal{L}(\theta, \lambda \mid Z, W, M) = \prod_{j=1}^J \prod_{i \in A_j} P_{ij} \quad (5.7)$$

Tanto los parámetros de pertenencia a las clases como los específicos a cada clase, deben ser estimados de manera simultánea, no es posible separar el problema. Esto es fácil de ver, pues no se tiene variable dependiente para estimar de forma aislada el modelo de pertenencia, ni tampoco se sabe a priori qué individuos pertenecen a qué clase como para estimar las preferencias de cada segmento de forma separada.

5.2 Estimación

El proceso de estimación de los parámetros sigue muy de cerca el trabajo de Bhat (1997), basado en maximización de función de verosimilitud (5.8), el que se extiende para solucionar un problema específico a la modelación realizada en este trabajo.

$$\begin{aligned} \begin{bmatrix} \hat{\theta} \\ \hat{\lambda} \end{bmatrix} &= \arg \max_{\theta, \lambda} \mathcal{L}(\theta, \lambda \mid Z, W, M) = \arg \max_{\theta, \lambda} \prod_{j=1}^J \prod_{i \in A_j} P_{ij} \\ &= \arg \max_{\theta, \lambda} \prod_{j=1}^J \prod_{i \in A_j} \sum_{s=1}^S P_{is} P_{ij|s} \end{aligned} \quad (5.8)$$

Bhat (1997) advierte que su estimación directa por máxima verosimilitud es bastante inestable, por lo que propone un proceso efectivo dividido en dos etapas⁸. Se inicia con el algoritmo de esperanza-maximización (*expectation-maximization algorithm*, Dempster,

⁷Bajo el supuesto de independencia entre las elecciones de los individuos (Greene y Hensher, 2003).

⁸Se intentó estimar de forma directa en este trabajo (proceso estándar), sin llegar a buenos resultados.

Laird, y Rubin, 1977) y es seguido por un método estándar de máxima verosimilitud cuando se encuentra “cerca” del óptimo (todos basados en algoritmo BHHH, Berndt, Hall, y Hall, 1974). Como se adelantó, se agrega una tercera etapa que no es más que una versión modificada de la segunda, que permite incluir variables que hacen del proceso de optimización inestable si se añaden al principio. El procedimiento fue programado en Matlab⁹, más detalles del proceso de estimación en anexo G.

⁹Inspirado en código `EMalgo.m`, implementación de Akshay Vij, estudiante de doctorado de University of California, Berkeley, para estimación de modelo *Latent Class* a través de *EM algorithm*. Se agradece su generosidad.

6. DISPONIBILIDAD Y CONSIDERACIÓN DE ALTERNATIVAS

La inclusión de alternativas relevantes para cada individuo es un aspecto de gran importancia en los modelos de elección discreta. En forma general, la teoría de la utilidad aleatoria está basada en un comportamiento compensatorio, el valor que asignan los individuos a una unidad marginal de los atributos de las alternativas permite que escasez de uno pueda ser compensado con más de otro¹. Se asume que las opciones son evaluadas por sus características (valoración hedónica), por lo que el modelo estima las preferencias por los atributos buscando el conjunto de parámetros que mejor se ajuste, dadas las elecciones y las no-elecciones².

La implicancia es que si se definen dentro del conjunto de alternativas algunas que no están realmente disponibles, por selección de parte de la oferta, imposibilidad de acceso, falta de información, u otros factores, los parámetros podrían resultar sesgados. De forma intuitiva, el sesgo se explica porque el modelo intenta representar la no elección de estas alternativas penalizando la valoración de los atributos más destacados de estas opciones³, pues no tiene como interiorizar que en realidad no están disponibles⁴.

En definitiva, al incluir alternativas no disponibles, los parámetros ya no representarían solo preferencias de los padres, sino que una función compleja de varios factores

¹Considerando dos atributos que son “deseables”.

²El modelo maximiza una utilidad que se puede escribir en función de la diferencia entre los atributos de la opción elegida y el resto de las alternativas disponibles no elegidas.

³En realidad, penalizando los atributos que más se diferencien respecto a las características de la opción elegida.

⁴En un ejemplo vinculado a este trabajo, no controlando por disponibilidad de colegios según selección académica, el modelo estima que un cuarto de las observaciones tienen valoración marginal por SIMCE negativa (prefieren un colegio con menor SIMCE a uno con mayor). Esto podría deberse a que muchas observaciones eligen un colegio de peor rendimiento existiendo mejores (con otras características iguales); cuando lo que sucede en la realidad es que estos establecimientos con mayor SIMCE no los tienen disponibles, por lo que aunque los prefieran, no tienen acceso y no pueden elegirlos.

mezclados con preferencia. Problema similar ocurre si existe dentro del proceso de elección mecanismos no compensatorios⁵, desviándose de los supuestos básicos de la modelación (Cantillo y Ortúzar, 2005). Diversos estudios han advertido el potencial sesgo del uso de un conjunto de alternativas inadecuado⁶ (Başar y Bhat, 2004; Swait y Ben-Akiva, 1986; Williams y Ortúzar, 1982), existiendo hoy opciones de modelación que permiten una definición endógena de éste⁷, reduciendo la probabilidad de error en estimación⁸.

En el contexto del estudio de elección de colegio, es tarea difícil definir el conjunto real de alternativas de las familias. Si existiera un sistema centralizado de postulación, que revele las alternativas consideradas (como en Hastings et al., 2007), o se realizara el gigantesco esfuerzo de preguntar directamente a los electores (como el estudio de Flores y Carrasco, 2013, pero a gran escala), se sabría el conjunto pertinente y dejaría de ser un problema. Pero para el general de los estudios sí lo es, y en vista de lo complejo del problema trabajos anteriores en Chile han optado por dejar todas las alternativas disponibles (Chumacero et al., 2011; Gallego y Hernando, 2008b). La excepción es Grau (2014), que define los establecimientos disponibles de cada observación como el conjunto de colegios elegidos por familias de la comuna de residencia, con similar nivel de educación de los padres.

Lo anterior no es el único problema, la eliminación de alternativas mediante procesos previos a la ponderación de atributos se confunde con la selección y la reputación de los colegios. Así como existen establecimientos no considerados porque no se conoce su

⁵La literatura vinculada con el proceso de elección destaca que los electores eliminan un gran número de alternativas en base a diferentes mecanismos, para quedarse con solo algunas pocas, las que evalúan con mayor detalle (Pagliara y Timmermans, 2009).

⁶Un conjunto de alternativas inadecuado contempla tanto alternativas no disponibles como alternativas que no se consideran en la elección, eliminadas por otros mecanismos distintos a la ponderación de sus atributos.

⁷Ver en Pagliara y Timmermans (2009) una revisión y comparación de diferentes propuestas. Otra alternativas no mencionadas revisar en de Lapparent (2010); Martínez, Aguila, y Hurtubia (2009).

⁸En este trabajo no se estiman tales modelos; el gran número de alternativas (colegios) hace prohibitivas la mayoría de las propuestas. Interesante tarea para futuros desarrollos.

existencia, o porque alguna característica lo descarta inmediatamente (comportamiento no-compensatorio), puede que algunos sean efectivamente preferidos, pero el acceso no es posible⁹. En otro orden, que la reputación del colegio en cuanto a selección (sea de cualquier tipo) implique su eliminación del abanico de alternativas, sin ser evaluado por otras características¹⁰.

El hecho de que el fenómeno de selección sea variable entre los colegios, y que dentro de cada colegio dependa de quién postula (características del alumno como de su familia), hace que la de modelación sea aún más compleja. Para los estudios nacionales, una práctica común para analizar la selección de los colegios es utilizar los requisitos del proceso de admisión declarados por los padres en el cuestionario SIMCE como criterio de selección¹¹ (Gallego y Hernando, 2008b; Grau, 2014; Madero y Madero, 2012), pero no se ha evaluado qué tan buen *proxy* es de la selección real. La aproximación más cercana es un estudio de correlación con prácticas reales declaradas por los directores (Carrasco et al., 2014). Como ya se ha mencionado, estas exigencias de matrícula podrían no significar en todos los casos selección, sino que diagnóstico o sencilla burocracia.

⁹En base a encuestas, Gallego, García, Lagos, y Stekel (2008) muestran que solo el 4 % de los padres declara haber sido rechazado, magnitud acorde con el 93 % de padres que declara en encuesta CEP 2006 estar en el colegio que quiere (Gallego y Hernando, 2008b). Aunque estos datos sugieren que la selección no es tan común, podría ser que no revelen de manera correcta el fenómeno, sino más bien éstos sean condicionales al conjunto de alternativas que ya no incluyen alternativas con reputación selectiva.

¹⁰En modelos no compensatorios, una de las posibilidades más interesantes es la “eliminación por aspectos”, descrita en Tversky (1972).

¹¹Una práctica que de alguna manera no respeta la base compensatoria de los modelos de elección discreta; se podría argumentar que en la mayoría de los casos la selección elimina la alternativa, no la hace “menos preferible”.

6.1 Construcción del Conjunto de Alternativas

Estudios en Chile confirman que los colegios realmente considerados por cada familia son muy pocos¹². Gallego et al. (2012) muestran que, en promedio, padres declaran que hay tres colegios en “el sector”. Schneider, Elacqua, y Buckley (2006) exhiben que el número de establecimientos considerados para matricular a los hijos es solo uno para el 53 % de las familias, dos para el 37 %, y tres o más para el resto. Flores y Carrasco (2013) llegan a valores muy similares: 55 %, 29 % y 16 % para las mismas cantidades anteriores. A partir de encuestas, Gallego et al. (2008) estiman que en la Región Metropolitana el promedio de colegios a los que postulan los padres es de 1,3. Si bien estos números son subjetivos, pues están relacionado con la etapa del proceso de elección en que se encuentran los padres, es bastante lejano al número de dos o tres cifras normalmente considerado en estudios basados en modelos de elección.

El presente trabajo no es la excepción, por razones ya mencionadas no es posible considerar el conjunto de alternativas real, pero sí intenta una reducción de las alternativas por medio de heurísticas, que buscan imitar un conjunto un poco más pertinente de opciones. La estrategia se basa en cinco criterios: distancia, ingreso, co-educación, etapa de elección y selección académica, todas con una fundamentación empírica.

6.1.1 Criterios de reducción de alternativas

Se consideró reducir alternativas por criterio de distancia bajo la hipótesis que existen “colegios de barrio”, en los que sus asistentes son en su mayoría vecinos, sobre los que familias que residen en zonas alejadas tendrían nula información o conocimiento. Lo anterior implica que se dejaría disponible, para cada individuo, solo los colegios que queden tan lejos como el que asiste actualmente, o colegios en que algún estudiante tiene que recorrer tanto como lo que él tendría que hacer para llegar.

¹²A pesar de que Raczynski et al. (2010) mencionan que “... los apoderados tienen un mapeo bastante completo sobre qué colegios existen en su barrio o cerca de sus hogares y saben de forma aproximada a cuántas cuadras de su casa quedan”.

El criterio de ingreso se basa en que existen colegios no considerados por las familias pues asisten individuos de nivel socioeconómico “distinto”. En concreto, se define como disponible un colegio si el ingreso del individuo está dentro del rango de ingresos de los asistentes a ese colegio¹³ o si la mensualidad del colegio es igual o menor a lo que declara pagar el individuo.

Co-educación es la restricción más simple, los colegios que atienden un solo género están disponibles únicamente para observaciones del sexo correspondiente. Etapa de elección se plantea como la no consideración de los colegios cuyo primer curso es séptimo (o niveles superiores), o colegios solo con enseñanza técnico-profesional, para las observaciones que estén matriculados en sexto en un colegio particular pagado y que no se cambian de colegio¹⁴.

Por último, la restricción de selección académica¹⁵ se basa en la “nota de presentación” de los individuos y la nota mínima requerida para el ingreso según colegio¹⁶. Se define que un colegio tiene procesos de selección académica, relacionados con el rendimiento anterior de sus postulantes, si la media de las notas de los alumnos que ingresan es mayor a la media de la nota del universo de alumnos que se cambian a colegios similares.

¹³Rango construido con un mínimo y máximo en base a distribuciones de ingreso de varias fuentes. Máximo se encontró solo para algunos colegios, dado que para bastantes casos era una variable truncada.

¹⁴Esto se basa en el supuesto que ya tienen definida su trayectoria educativa desde antes (primeros años escolares).

¹⁵La importancia de la modelación de la selección académica surge de repetidas estimaciones inconsistentes de la valoración por SIMCE. Sin incluir este criterio, un cuarto de las observaciones valora negativamente un aumento en SIMCE del colegio. Se cree que esa es la manera en que el modelo internaliza que estén disponibles colegios de bajo cobro y alto rendimiento, y que no son frecuentemente elegidos, cuando en realidad puede que estos colegios no estén disponibles para gran parte de la población debido a selección académica.

¹⁶En la práctica la nota mínima no es el único requisito. Se establece un límite, ligado a la moda de la distribución de notas de presentación de las observaciones que ingresan realmente al colegio. Observaciones con nota superior o igual tienen disponible el colegio. Para individuos con nota menor a este límite y mayor a la nota mínima de la distribución anterior, se asigna una “probabilidad de disponibilidad”, en base a una distribución empírica. Más detalles en anexo E.

En el anexo F se analizan las notas de presentación de los alumnos según colegio, clara evidencia que los colegios con resultados destacados reciben alumnos con mejor rendimiento medido según nota del colegio anterior. En el mismo anexo se define con detalle qué colegios se consideran dentro de los que seleccionan, junto con un pequeño análisis.

Los detalles de definición y construcción de las restricciones del conjunto de alternativas se desarrollan en el anexo E, junto con un análisis del efecto de éstas en la disponibilidad geográfica de colegios.

6.2 Conjunto Final de Alternativas

El conjunto de criterios que limitan la disponibilidad de alternativas se traduce en una reducción del número de colegios disponible promedio por observación, de 798 (todos los establecimientos) a 189. Los criterios mencionados anteriormente concuerdan con los resultados de Raczynski et al. (2010), que postula que en una primera etapa del proceso de elección los padres excluyen alternativas en base a precio y distancia. En la Tabla 6.1 se muestra el resultado de disponibilidad de colegios según decil de ingreso luego de imponer las restricciones mencionadas.

TABLA 6.1. Cantidad de colegios disponibles promedio según decil y tipo de restricción

Decil	Restricción Ingreso			Restricción Distancia			Restricción Notas			Todas		
	Muni	Subv	Part	Muni	Subv	Part	Muni	Subv	Part	Muni	Subv	Part
(1)	153	332	15	65	154	40	132	392	151	46	91	4
(2)	156	419	20	64	156	38	134	400	154	48	115	5
(3)	156	438	25	64	156	38	135	405	156	48	120	6
(4)	156	456	37	65	158	41	135	404	155	49	126	9
(5)	156	463	55	65	159	42	137	408	157	49	130	14
(6)	156	464	79	67	163	45	137	410	158	51	133	21
(7)	155	462	108	68	163	48	138	414	159	51	132	28
(8)	151	449	152	69	163	55	139	418	161	49	128	41
(9)	118	390	178	70	148	76	139	427	164	35	98	62
(10)	64	280	177	60	99	99	142	445	172	14	50	83

La restricción de ingreso claramente afecta la disponibilidad de colegios para los deciles más bajos, en especial a través de la disminución de colegios privados (177 a 26 para

el primer decil), y de subvencionados (464 a 381 para el primer decil). Cabe destacar que la reducción de alternativas subvencionadas es de gran magnitud (sobre un 15 %) solo para el decil más pobre y los dos más ricos, mientras que la reducción en colegios privados es relevante en los 7 primeros deciles. Las opciones municipales solo disminuyen para el grupo más rico: en más de la mitad no asiste nadie con ingresos mayores o iguales al percentil 90 (efecto similar pero en menor magnitud ocurre con los subvencionados).

Cuando se restringe por distancia se observa que todos tienen acceso a una cantidad similar de colegios municipales. Las diferencias aparecen en la disponibilidad de colegios subvencionados: los nueve primeros deciles tienen casi igual número de alternativas, mientras que el decil más rico tiene un tercio menos de colegios disponibles. Efecto contrario sucede con los colegios privados, deciles del 1 al 7 tienen acceso similar, aumentando en un 30 %, 80 % y 135 % para los deciles 8, 9 y 10 respectivamente. Este fenómeno se puede atribuir a una oferta de colegios más caros condicionada a la segregación espacial de los deciles más ricos. Basta observar la Figura 6.2, que muestra la distribución geográfica de la población según decil de ingreso; y la Figura 6.1, que ubica los colegios según dependencia, para notar que el resultado anterior es esperable.

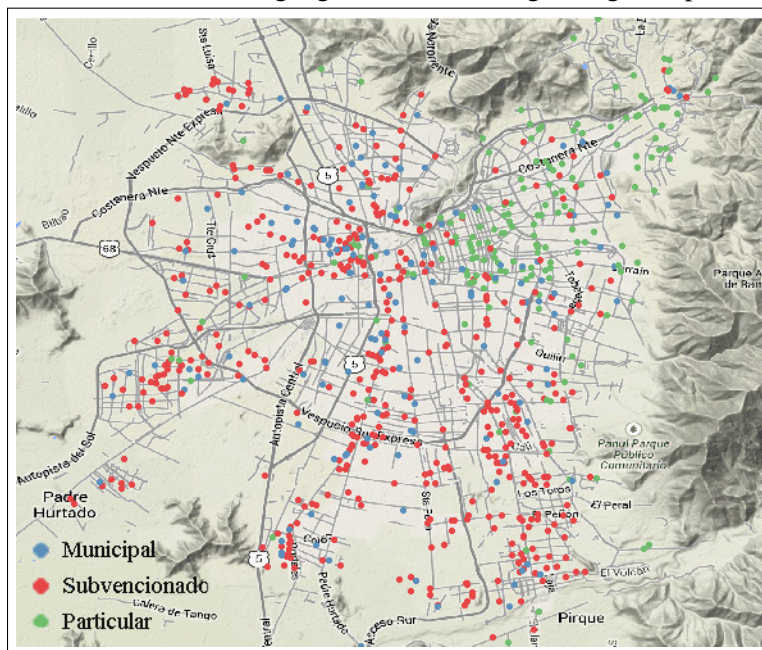
La restricción académica afecta de forma diferenciada según grupos. Individuos con mayor ingreso tienen mejor disponibilidad, lo que, por construcción, está ligado únicamente a mejores notas de presentación¹⁷.

Todas las restricciones en su conjunto limitan severamente la oferta de colegios privados desde el noveno decil hacia abajo, reduciendo la cantidad desde un 25 % para el noveno decil hasta un 93 % para el primer decil si se compara con el grupo de mayor ingreso. También restringen la oferta de colegios subvencionados para el quintil más rico, por ubicación geográfica e ingreso elevado, y para el primer decil, por restricción de ingreso. Es importante recordar que los datos son previos a la ley de subvención escolar

¹⁷El dato de disponibilidad para los dos deciles más afluentes no se debe solo a mejores notas. La disponibilidad para observaciones de establecimientos privados que no se cambian de colegio no se restringió, pues se supuso que su elección se produce al iniciar el proceso escolar del alumno, donde no exigen notas de colegios anteriores. Ver detalles en anexo E.

preferencial, que puede haber ampliado la disponibilidad de colegios para los grupos más vulnerables.

FIGURA 6.1. Ubicación geográfica de los colegios según dependencia



[illegible]

7. RESULTADOS MODELO DE ELECCIÓN

En este capítulo se presentan y discuten los resultados de estimar el modelo de elección de colegio descrito anteriormente, que considera las restricciones en disponibilidad. En la línea de otros trabajos basados en la misma metodología (Walker y Li, 2007; Wen y Lai, 2010), se evalúa el ajuste del modelo según el BIC (criterio de información bayesiano) para determinar el número de clases. Finalmente se eligió un modelo logit de clases latentes con cinco segmentos, cuyos resultados se resumen a continuación¹.

Es necesario advertir que para la estimación se utiliza el colegio finalmente escogido por las familias, y no el colegio preferido. Estos podrían ser distintos debido a que los colegios también eligen a través de los procesos de selección, o por falta de información². El resultado es que la elección final podría no ser una fidedigna realización de las preferencias. A pesar de las diversas restricciones que se consideraron para intentar dejar el conjunto de alternativas que realmente disponibles, es difícil que reflejen fielmente los diversos métodos de selección que existen en el sistema escolar, y la subjetividad de éstos.

¹Resultados correspondientes al archivo `resStage1.1hr9.11-5.mat`

²Se entiende como “colegio preferido” el establecimiento que, dada la restricción de presupuesto de la familia, la capacidad y disposición al viaje y todos los otros atributos por los que se controla en la modelación, es el que maximiza la utilidad de la familia. Podría diferir del colegio elegido por elementos no considerados, como selección por valores de la familias, religión, tradición familiar, etnia o cualquier restricción ausente o incorrectamente considerada en la modelación. Si se incluyeran realmente todas las restricciones de disponibilidad (selección), y se conociera el nivel de información real de las familias, entonces no debería existir diferencia entre el colegio elegido y el preferido (dentro del conjunto de colegios disponibles y conocidos)

7.1 Clases

El modelo define cinco clases o tipos de demandante por colegio diferenciados fuertemente por el nivel socioeconómico de la familia: ingreso y educación³. Como se observa en la Tabla 7.1, cada segmento es superior en todos sus atributos al segmento anterior, lo que tiene sentido dada la alta correlación entre las características⁴. De la misma tabla se observa que los parámetros de pertenencia son todos significativamente distintos a cero, a excepción de la clase 3 (clase base), cuyos parámetros se fijaron en cero para asegurar identificación.

Las características medias de las clases uno y dos corresponden a padres sin educación media terminada, con bajos ingresos; aunque los de la clase dos tienen mayor escolaridad y recursos. El segmento tres representa apoderados con educación media completa, estudios superiores de corta duración (un año), e ingresos cercanos a las mediana de la población. El cuarto son padres con cerca de tres años de estudios terciarios promedio, e ingresos iguales al percentil 70. Por último, la clase cinco consideraría padres con estudios universitarios completos e ingresos que representan el 20 % más rico.

7.2 Preferencias (según clase)

Se estimaron tantos conjuntos de parámetros como el número clases escogido. Siguiendo el trabajo de Gallego y Hernando (2008b), se permitió que la ponderación de características de los colegios variara de acuerdo a atributos individuales⁵. Los resultados de estimación se muestran en las Tablas 7.2, 7.3 y 7.4. Las columnas de la segunda a la sexta representa los parámetros estimados para cada clase.

³Las características de segmentación fueron logaritmo del ingreso per cápita (logIncomePC), años de educación de la madre (edMotherAnhos), años de educación del padre (edFatherAnhos), únicas variables continuas de la base. No tiene mucho sentido agregar variables binarias a la segmentación.

⁴corr(logIncome,eduMother)= 0, 60; corr(logIncome,eduFather)= 0, 63;
corr(eduMother,eduFather)= 0, 65.

⁵Respecto a los atributos no considerados para la segmentación, que comprenden solo variables binarias.

TABLA 7.1. Características promedio de individuos según clase

Clase	Clase				
	1	2	3	4	5
Características					
Porte Segmento	0.36	0.09	0.27	0.09	0.18
Miembros	15012	2499	10783	2493	7072
Educación de la madre [años]	9.69	10.44	13.76	13.99	16.19
Educación del padre [años]	9.88	10.61	13.88	14.39	16.80
Ingres [log(CLP)]	10.43	10.70	11.43	11.56	12.91
Parámetros del logit condicional (modelo de pertenencia a clases)					
cte	-0.944	-1.739	-	-1.197	-4.822
Educación de la madre	-1.565	-1.389	-	-0.014	1.068
Educación del padre	-1.268	-1.093	-	0.166	0.849
Log(Ingreso)	-2.276	-1.783	-	0.266	3.763

Todos los parámetros presentan un valor-p < ,01.

326 de los 367 parámetros finalmente incluidos son estadísticamente diferente de cero (74 % con significancia al 1 %, 13 % al 5 % y 5 % al 10 %). El modelo es informativo y pasa los test tradicionales de ajuste (Ortúzar y Willumsen, 2011), (ver detalles en sección G.6 del anexo G).

La metodología utilizada permite distinguir patrones de elección por clase, pero dada la gran cantidad de interacciones incluidas es difícil realizar interpretaciones directas desde de los parámetros del modelo. Para un análisis general es más informativo revisar los efectos marginales, calculados en la siguiente sección. Al agruparlos por decil, aprovechan la estructura de la estimación de varios segmentos y permiten observar la heterogeneidad según nivel de ingreso. Sí es interesante verificar consistencias básicas y algunas hipótesis directamente de los parámetros de estimación.

Siguiendo el orden de las Tablas 7.2 y 7.3 se aprecia que padres con expectativas universitarias de sus hijos buscan colegios con mejor rendimiento, tienden a preferir colegios más grandes, sin currículum técnico profesional, y, específicamente los de menor ingreso (clase uno), están dispuestos a viajar más y a asistir a colegios más caros. Entre los apoderados para los que la formación valórica es prioridad, son preferidos los establecimientos

católicos, con requisitos ligados a ritos religiosos, con mejor calidad percibida en entrega de valores, mientras que el rendimiento les importa menos.

Familias que consideran relevante el prestigio en la elección, prefieren colegios con mejor SIMCE, con mayores exigencias en la admisión, y en general son más sensibles al cobro. Padres que declaran importante la cercanía en la elección efectivamente les importa más la distancia y menos el nivel de ingreso de los vecinos, mientras que los que declaran importante el nivel socioeconómico son menos sensibles al copago. Familias que dicen valorar el puntaje SIMCE o la PSU en la elección prefieren colegios con mejor rendimiento SIMCE, y además les es menos relevante la calidad percibida en entrega de valores. Aquellos que eligen porque se adecua al presupuesto familiar penalizan más fuerte el copago. Por último, entre los apoderados que eligen colegio por tradición familiar, en promedio prefieren colegios católicos y de mayor tamaño.

7.3 Efectos Marginales

Se calculó el efecto marginal del cambio en atributos del colegio sobre la probabilidad de elección del mismo⁶. Su interpretación es el cambio de la probabilidad de elección del colegio real (en puntos porcentuales), frente a un incremento de una desviación estándar en un atributo del colegio (en caso de que sea continuo). Vale la pena reiterar que es válido “solo en el margen”.

Dada la heterogeneidad en preferencias, y su dependencia con las características de los individuos, un análisis pertinente es el comportamiento marginal según decil de ingreso per cápita (Tabla 7.6). En anexo el H se exhiben Figuras que representan gráficamente la variabilidad que existe.

Los efectos marginales promedio por decil de ingreso (Tabla 7.6) muestran que hay clara diferencia según el nivel de recursos en la ponderación de atributos relevantes en la elección. Los cuatro atributos más importantes, dentro de los incluidos, son la distancia,

⁶Se calcularon a nivel individual y promediaron de dos diferentes formas. Para detalles de cálculo ver sección G.5 en anexos.

el copago, el puntaje SIMCE y el tipo de enseñanza. La distancia es una característica valorada negativamente por todos, pero casi el doble en magnitud de relevancia para el grupo más pobre, en comparación con el más rico.

El copago disminuye fuertemente la probabilidad de elección para el segmento más vulnerable, bajando su incidencia en la decisión a medida que aumenta el ingreso hasta el noveno decil. Para el 10 % con más recursos, el modelo estima una relación positiva entre probabilidad de elección y mayor copago⁷.

El tipo de enseñanza es un atributo con bastante diferencia según nivel de ingreso. Hasta el tercer decil, en promedio las familias prefieren un colegio con currículum técnico profesional, mientras que desde el sexto decil en adelante es un atributo que disminuye la probabilidad de elección del colegio. Los colegios con enseñanza científico-humanista son valorados por todos, con magnitud creciente según ingreso.

Se observa que mayor puntaje SIMCE aumenta la probabilidad de elección para todos, pero con una ponderación más del doble si se compara la del octavo decil con la del primero. Los resultados muestran que el *peak* de valoración está en el octavo decil, el 20 % con más recursos tiene preferencias por SIMCE cercanas a las del sexto decil.

Siguiendo a Chumacero et al. (2011), se calcularon las tasas marginales de sustitución⁸ relacionadas con el SIMCE, distancia y cobro. En vista que la metodología permite el cálculo a nivel de individuo, en la Figura 7.1 se muestra la mediana por decil de ingreso⁹. Se observa que los más pobres están dispuestos a viajar considerablemente menos por

⁷Si bien puede ser un problema de endogeneidad por variables omitidas, parte del efecto podría ser entendido con la hipótesis postulada en Raczynski et al. (2010) sobre la relación entre copago y estatus, o la relación directa que hacen los padres entre mayor pago y atributos valorados (en el sentido que para ciertos padres podría ser un *proxy* de calidad).

⁸Compensaciones frente a cambios en atributos que mantienen igual probabilidad de elección del colegio.

⁹Solo se exhiben los resultados hasta el octavo decil para el caso que involucra cobro. Esto se debe a que la tasa de sustitución relacionada con pago carece de interpretación para el 20 % más afluente, pues el efecto marginal de un mayor copago se estimó positivo para los más ricos.

un colegio con mejor rendimiento: solo 40 metros por un incremento en 10 puntos; mientras que los grupos más ricos recorrerían sobre 160 metros por los mismos puntos. Efecto similar ocurre con el pago, la mediana de la disposición del primer decil son \$3.500, y del octavo decil \$23.000 por igual aumento en rendimiento. Respecto a la sustitución pago-distancia, el grupo más vulnerable pagaría \$750 por recorrer 10 metros menos, mientras que el octavo decil \$1.450.

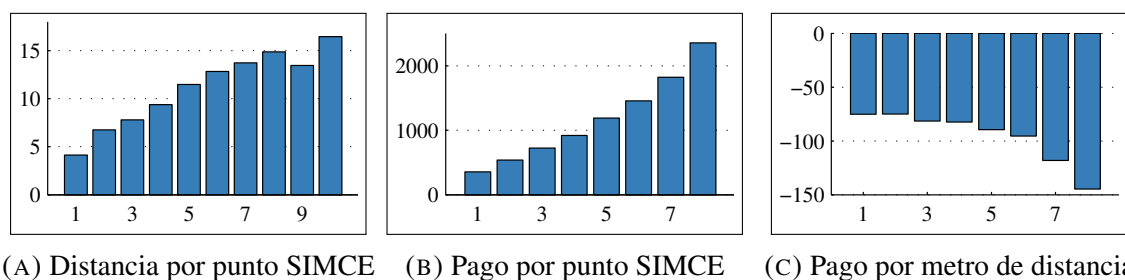


FIGURA 7.1. Medianas de tasas marginales de sustitución (eje Y) según decil de ingreso (eje X).

Volviendo a los efectos marginales, el tamaño del colegio, medido como número de alumnos en educación media, es valorado de forma positiva por todos. Colegios con examen son preferidos con magnitud creciente hasta el quinto decil¹⁰, donde experimentan una baja, y no siendo relevantes para el 20 % de mayores recursos. La presencia de entrevista, en cambio, es un atributo no deseado para el 60 % con menores recursos, decreciente en magnitud al aumentar el ingreso, y transformándose en característica atractiva para el 30 % más afluente. Mismo efecto sucede con la preferencia por colegios católicos, pero el cambio a positivo es en el séptimo decil.

Los atributos restantes no presentan comportamiento monótono según nivel de ingreso. Hasta el octavo decil un colegio en un barrio con mejores ingresos aumenta la probabilidad de elección del colegio, no siendo importante en la elección para el sector con más

¹⁰Raczynski et al. (2010) encuentran que, en general, familias de estratos bajos y medios valoran los procesos de selección, pues les aseguran un mejor clima escolar y “evita el contacto de sus hijos con estudiantes que tienen otros valores y conductas indeseadas a sus ojos”.

ingreso. La infraestructura es una característica valorada de forma transversal, aumentando en magnitud en los sectores medios. Por último el modelo estima que solo el 30 % más vulnerable prefiere colegio que entreguen buena formación valórica¹¹, mientras que para resto es una característica que disminuye la probabilidad de elección (salvo en el cuarto y décimo decil, para los que, en promedio, no es relevante).

7.4 Reflexiones a partir de los resultados

Si las estimaciones realizadas reflejaran profundas preferencias de los padres, entonces la segregación del sistema educacional (en términos de ingreso) sería difícil de superar únicamente haciendo más accesible la oferta. Por más que el sistema no tenga restricciones a través de pago o selección, la alta ponderación que le entregan las familias con menos recursos a la distancia, sea por restricciones de movilidad, seguridad, tiempo disponible, costo u otras variables, no permitiría que se encuentren en los establecimientos con familias de mayor capital cultural, quienes están dispuestos a viajar más (sin contar la segregación residencial, que agrava el problema). Se agrega la diferencia en importancia respecto al desempeño académico, padres con más recursos perciben un mayor retorno al encontrar un mejor colegio, lo que desequilibra las motivaciones de búsqueda. En este escenario, los caminos posibles hacia la integración serían dos: cambiar las preferencias de las personas o limitar la libre elección para forzar integración, estrategias impracticables o, al menos, políticamente discutibles.

¹¹Recordar que este atributo, al igual que infraestructura, está basado en la percepción de los padres sobre su propio colegio, no en medidas objetivas o criterios transversales.

El otro escenario, y que parece ser razonable, es que los parámetros reflejen una combinación entre preferencias, nivel de información¹², restricciones y selección no consideradas, y otros factores¹³. En este caso, las propuestas para generar un sistema más integrado son menos evidentes. Si bien es necesario tener una oferta más accesible, sea eliminando el copago, prohibiendo y fiscalizando la selección, creando programas de transporte inteligentes (o soñar una ciudad más integrada), u otras políticas; cualquier estrategia debe ser acompañado de mayor información que repercuta en mayor empoderamiento de los padres en el proceso, y una elección de “mejor calidad”, por sobretodo en los sectores de mayor vulnerabilidad.

¹²La teoría en que se basa el modelo elegido asume información perfecta, supuesto que claramente no se cumple y, aún más complicado, con grado de conocimiento dispar entre observaciones, y seguramente correlacionado con el nivel socioeconómico (Raczynski et al., 2011). Azmat y Garcia Montalvo (2012), con datos de Pakistán, encuentran gran disparidad en precisión de la información en el contexto de elección de colegio según educación de la madres, e importante impacto de ésta en la decisión.

¹³No se descarta sesgo debido a endogeneidad por variables omitidas. Solo por mencionar algunas, no se incluye en la estimación principal las características socioeconómicas del alumnado de los colegios, aspecto que puede ser relevante para los padres.

TABLA 7.2. Tabla de parámetros modelo Logit Clase Latente, A

Variable nivel colegio (Variable nivel estudiante en negrita)					
Clase	(1)	(2)	(3)	(4)	(5)
Constante					
Sqrt. Distance	-1.224**	-0.803**	-1.008**	-1.301**	-0.933**
SIMCE	-0.022	1.163**	0.878**	0.316*	-0.550**
Requisito examen	0.258**	-1.908**	0.909**	-1.510**	-0.174
Requisito entrevista	-0.088	-0.486**	0.259**	-1.027**	0.609**
Requisito religioso	-0.107	-1.890**	-0.001	-0.942**	-0.040
Formación valórica	0.026	-0.019	-0.321**	-0.182**	-0.014
Infraestructura	0.030	-0.249**	0.109*	0.431**	0.008
Nombre en inglés	-0.268 ⁺	0.111	-0.715**	0.445**	-0.089
Católico	-0.922**	0.888**	-0.724**	0.429*	-0.014
Tamaño	0.305**	0.166**	0.318**	0.470**	0.316**
Técnico-Profesional	1.171**	4.439**	0.297*	-0.199	-1.163**
Científico-Humanista	-0.375**	-0.206 ⁺	-2.144**	2.950**	1.181**
Ingreso barrio	0.168**	-0.404**	0.180**	0.093	-0.137 ⁺
Mensualidad	-1.531**	-0.267	-1.709**	-0.139	0.119
Mujer					
SIMCE	-	-0.176 ⁺	-	-	0.248**
Requisito examen	-0.220**	-0.424*	-0.173*	-	-
Requisito entrevista	-0.389**	-	-	0.603**	-0.302**
Requisito religioso	-0.554**	2.848**	-	-	-0.358**
Formación valórica	-	-	-0.172**	-	-
Infraestructura	-	-	-0.108**	-	-
Nombre en inglés	0.212 ⁺	-	-	-	-
Católico	0.608**	-	0.643**	-	0.107
Tamaño	0.069**	0.107**	-	-	-
Técnico-Profesional	-0.173**	-	-0.320**	-	-
Científico-Humanista	-	-	0.265*	-	-
Ingreso barrio	0.102**	0.458**	-	-	-
Mensualidad	-	-	-	-0.245*	-
Expect. Ed. técnica					
Sqrt. Distance	0.099*	-	-	-	-
Requisito examen	0.257**	-	-	-	-
Católico	-	-	-	-	-1.342*
Científico-Humanista	-	-	0.428*	-	-
Mensualidad	0.826**	-	-	-	-
Expect. Ed. universitaria					
Sqrt. Distance	0.194**	-	-	-	-
SIMCE	0.377**	0.449**	-	0.361**	0.777**
Requisito examen	0.225**	-	-	-	-
Católico	0.166*	-	-	0.058	-1.228*
Tamaño	0.053**	-	0.071**	0.263**	-
Técnico-Profesional	-0.423**	-0.950**	-0.321*	-3.763**	-
Científico-Humanista	0.279**	-	1.465**	-	-
Mensualidad	0.790**	-	0.615**	-	-

** valor-p < 0,01; * valor-p < 0,05; ⁺ valor-p < 0,1,

TABLA 7.3. Tabla de parámetros modelo Logit Clase Latente, B

Variable nivel colegio (Variable nivel estudiante en negrita)					
Clase	(1)	(2)	(3)	(4)	(5)
Pref. valores					
Sqrt. Distance	-	-0.192*	-0.221**	-	-0.113**
SIMCE	-0.096*	-	-0.425**	-	-0.229**
Requisito examen	0.231**	-1.483**	-	-	-0.223*
Requisito entrevista	-	0.956**	-	1.398**	-
Requisito religioso	-	1.010**	-	2.213**	0.271**
Formación valórica	0.554**	0.984**	0.729**	-	1.000**
Infraestructura	-0.080*	-	-	-	-0.151**
Católico	0.874**	-	1.373**	-	0.932**
Tamaño	-	-	-	-0.080	-
Técnico-Profesional	-	-	-0.589**	-	-
Científico-Humanista	-0.086	0.358*	-	-	-
Ingreso barrio	-	-0.444**	-	-0.402**	-
Mensualidad	0.524**	-	0.405**	-	-
Pref. prestigio					
Sqrt. Distance	-	-0.226**	0.156**	-	-
SIMCE	0.878**	0.483**	0.803**	0.782**	0.919**
Requisito examen	0.495**	-	0.422**	-	-
Requisito entrevista	-	-	-	-	0.221*
Requisito religioso	-	0.437*	0.211 ⁺	-	0.173*
Formación valórica	0.117**	0.647**	-	-	-0.140**
Infraestructura	0.215**	-0.229**	0.207**	-	-
Nombre en inglés	-	-	-	-	0.126 ⁺
Católico	-0.685**	-	-0.492**	0.533**	-
Tamaño	-	0.110**	-	-	0.323**
Técnico-Profesional	-	-	0.391**	-	-
Mensualidad	-0.827**	-	-0.623**	-0.612**	-
Pref. proximidad					
Sqrt. Distance	-2.231**	-0.241 ⁺	-2.222**	-0.278	-1.248**
SIMCE	0.201**	-1.059**	0.130 ⁺	-0.315	-
Requisito examen	-0.411**	1.866**	-0.501**	0.614	-
Requisito entrevista	-0.353**	0.479*	-0.499**	-	0.282**
Requisito religioso	-	-1.099**	-	-	-
Formación valórica	-	-	-	-	-0.253**
Infraestructura	-	0.406**	0.093*	-	-
Nombre en inglés	0.367**	-	0.635**	-	0.334**
Católico	0.208*	-1.345**	-	-	0.403**
Tamaño	-	-	0.060**	-	-
Técnico-Profesional	-	-2.082**	-	-	-
Científico-Humanista	0.414**	0.495**	0.905**	-	-
Ingreso barrio	-0.197**	-	-0.185**	-	-
Mensualidad	-	-	0.282**	-	-0.117**

** valor-p < 0,01; * valor-p < 0,05; ⁺ valor-p < 0,1,

TABLA 7.4. Tabla de parámetros modelo Logit Clase Latente, C

Variable nivel colegio (Variable nivel estudiante en negrita)					
Clase	(1)	(2)	(3)	(4)	(5)
Pref. nivel socioeconómico					
Sqrt. Distance	0.153*	-0.688**	-0.243**	-	-
Requisito examen	0.329*	-1.499**	-	-	-
Requisito entrevista	-	1.200**	-	-	-
Requisito religioso	-	-	0.557*	-2.250**	-
Infraestructura	-	-	0.170*	-	0.114*
Nombre en inglés	0.645**	-	0.675**	-	-
Católico	-0.430*	-	-0.528**	-	-
Técnico-Profesional	-	-1.639**	-	-	-
Científico-Humanista	-	-	2.204**	-	-
Ingreso barrio	-	-	-0.125*	-	-
Mensualidad	0.893**	-	0.487**	0.760**	0.211**
Pref. SIMCE					
Sqrt. Distance	-	-0.553**	-	-	-0.261**
SIMCE	0.437**	1.115**	0.644**	0.450**	0.765**
Requisito examen	-0.172	-	-0.435**	-	-
Requisito entrevista	-	-1.511**	-	-	-
Infraestructura	0.351**	-	-	0.463*	-
Nombre en inglés	-	0.757 ⁺	-	-	-
Católico	-0.530**	-	-0.935**	-	-
Tamaño	-	-	-0.213**	-	-
Científico-Humanista	-	-	0.867**	-	-
Mensualidad	-	-	-0.902**	-0.942**	-0.333**
Pref. PSU					
Sqrt. Distance	-	-	-	0.697**	-
SIMCE	0.486**	0.630*	1.425**	0.833**	1.775**
Requisito examen	0.909**	1.255**	-0.811**	2.130*	-
Requisito entrevista	1.295**	-	1.143**	-	-
Requisito religioso	2.167**	-3.182**	-	-	-
Formación valórica	-0.706**	-0.964**	-	-1.864**	-0.501**
Infraestructura	0.498**	0.704**	-0.641**	1.756**	-0.412**
Católico	-1.894**	-	-	-	0.504**
Tamaño	0.193**	-	0.407**	0.271**	0.356**
Técnico-Profesional	-0.764**	-2.077**	-	-	-
Científico-Humanista	2.230**	-	-	-	-
Ingreso barrio	-	-0.252	-	-	0.142**
Mensualidad	-2.621**	-	-	-	-

** valor-p < 0,01; * valor-p < 0,05; ⁺ valor-p < 0,1,

TABLA 7.5. Tabla de parámetros modelo Logit Clase Latente, D

Variable nivel colegio					
(Variable nivel estudiante en negrita)					
Clase	(1)	(2)	(3)	(4)	(5)
Pref. costo					
Sqrt. Distance	-	-	0.173**	-	-0.153**
SIMCE	-	-	-	-	0.190**
Requisito examen	-	0.415**	-	-	-
Requisito entrevista	0.146 ⁺	-0.553**	-	-	-
Requisito religioso	-	-0.675**	-	0.491*	-0.281**
Formación valórica	-0.163**	-	-	-	-0.192**
Nombre en inglés	-	-	0.401**	-	-
Católico	0.271**	-	-	-	0.450**
Tamaño	0.087**	-	-	-	-
Técnico-Profesional	-0.643**	-0.724**	-	-	-
Científico-Humanista	0.134*	-	0.776**	-	-
Ingreso barrio	0.161**	-	0.051	-	-0.120**
Mensualidad	-4.439**	-1.516**	-1.039**	-0.857**	-0.324**
Pref. tradición					
Sqrt. Distance	-0.079*	-	-0.181**	-	-
SIMCE	-	0.199*	-	-	0.429**
Requisito examen	-	-	-0.071	0.372 ⁺	-
Requisito entrevista	-	-	-	-	-0.512**
Formación valórica	-	-	-	-	-0.126**
Infraestructura	-0.076**	-	-	-	0.118**
Nombre en inglés	-0.336*	-	-	-	-
Católico	0.204*	-	0.746**	-	0.351**
Tamaño	-	-	0.114**	-	0.163**
Científico-Humanista	-	-0.509**	0.877**	-	-
Ingreso barrio	-	-	0.147**	-	-
Mensualidad	-	-	-	-	0.126**
Auto					
Sqrt. Distance	-	-	-	-	0.115
SIMCE	-	-	-	0.468**	-
Requisito examen	0.295**	0.205	-	-0.587**	0.334 ⁺
Requisito religioso	-	0.359 ⁺	-	-	-
Formación valórica	-	-	-	-	-0.334**
Infraestructura	-	-	-	-	0.221**
Nombre en inglés	-	-	0.235 ⁺	-	-
Católico	-	-	-	-	0.626**
Técnico-Profesional	-0.280**	-0.393*	-	-	-0.798 ⁺
Ingreso barrio	-	-	-	-	0.160*
Mensualidad	-	-	0.677**	0.383**	0.283**

** valor-p < 0,01; * valor-p < 0,05; ⁺ valor-p < 0,1,

TABLA 7.6. Efectos marginales de cambio atributos de colegio por decil de ingreso

	Decil									
	1	2	3	4	5	6	7	8	9	10
Raiz de Distancia	-17.61**	-15.33**	-14.62**	-13.45**	-12.86**	-12.78**	-12.22**	-11.58**	-10.60**	-9.45**
Mensualidad	-19.85**	-16.82**	-15.65**	-13.56**	-12.46**	-11.16**	-8.38**	-5.88**	-0.90**	2.04**
SIMCE	3.95**	4.62**	5.14**	5.52**	6.58**	7.56**	8.50**	9.36**	7.75**	7.24**
Católico	-1.18**	-1.03**	-0.93**	-0.46**	-0.44**	-0.24	0.05	0.38 ⁺	1.65**	2.84**
Tamaño	2.70**	2.54**	2.55**	2.52**	2.66**	2.85**	3.07**	3.35**	3.30**	3.71**
Formación valórica	0.64**	0.57**	0.37**	0.05	-0.35**	-0.73**	-1.19**	-1.71**	-1.37**	-0.21
Ingreso Barrio	0.80**	0.63**	0.58**	0.53**	0.51**	0.48**	0.42**	0.37**	0.11 ⁺	0.13
Intraestructura	0.61**	0.70**	0.83**	0.89**	1.09**	1.28**	1.49**	1.56**	1.02**	0.87**
English Name	0.42	0.37	0.40	0.51*	0.50*	0.57*	0.85**	1.01**	1.13**	0.95**
Curriculum										
Cient.-Humanista	0.79**	1.04**	1.26**	1.58**	2.08**	2.56**	3.19**	4.15**	4.60**	5.02**
Téc.-Profesional	3.82**	3.04**	2.52**	1.52**	0.55**	-0.29	-1.52**	-3.05**	-5.16**	-6.58**
Entry requirements										
Interview	-2.40**	-1.75**	-1.43**	-1.06**	-0.51**	-0.17	0.30 ⁺	0.94**	2.07**	2.69**
Test	0.74**	0.88**	0.88**	0.88**	0.96**	0.90**	0.64*	0.24	-0.04	0.00
Religious	-1.64**	-0.97**	-0.68*	-0.38	0.06	0.28	0.54*	0.36 ⁺	-0.05	0.12

** valor-p < 0,01; * valor-p < 0,05; + valor-p < 0,1,

8. EJERCICIOS

Con el fin de cuantificar los efectos de diferentes características que definen el sistema de educación media sobre la segregación socioeconómica, se procedió a realizar ejercicios de equilibrio parcial. Utilizando el modelo previamente estimado, se calculó la segregación que resulta de la hipotética elección de los padres en escenarios con cambios respecto al real (contrafactuales). No es la primera vez que se realizan aproximaciones como ésta, Gallego y Hernando (2008a) desarrollan un ejercicio con el que calculan cambios en segregación¹, Neilson (2013) estima los mecanismos inducidos por la subvención escolar preferencial que explican el impacto de ésta en la población más pobre, y Grau (2014) evalúa cambios en matrícula frente *shocks* exógenos en SIMCE². Si bien lo óptimo sería utilizar un modelo que permita estimar un equilibrio general en base a los escenarios propuestos, se aleja de las ambiciones del trabajo, sin dejar de pretender ser informativo en el contexto del estudio empírico de la segregación.

8.1 Diseño

Los ejercicios realizados fueron simples en construcción, algunos bastantes alejados de lo que podría ser una política pública viable, y todos carentes de un equilibrio general³.

¹El énfasis del trabajo está más bien puesto en cuantificar el beneficio de un sistema de libre elección que en segregación.

²Los últimos dos ejemplos utilizan modelos estructurales. Algunos trabajos internacionales que usan contrafactuales son Burgess, McConnell, Propper, y Wilson (2007); Lindbom (2010); Östh, Andersson, y Malmberg (2013).

³Solo por mencionar, algunos fenómenos que no están siendo tomados en cuenta son: la apertura y cierre de colegios, aumentos o disminución de cupos, cambios en atributos de colegios en respuesta a modificaciones en la demanda o imposiciones en la oferta. Posiblemente, y dado que los parámetros no representan de forma pura las preferencias de las personas, también los parámetros estimados podrían cambiar por efectos en la oferta no considerados.

Con el modelo estimado anteriormente, se calculó la probabilidad de elección de los colegios disponibles para cada observación, previo a introducir cambios en características de los colegios y/o en las familias, así como modificaciones en las preferencias⁴.

Se definió el colegio predicho por el modelo a partir de los tres colegios con mayor probabilidad⁵. Dentro de éstos se eligió de forma aleatoria, con ponderadores proporcionales a la probabilidad predicha por el modelo⁶.

Dado que cada colegio tiene cierta capacidad, se asumió una cantidad de cupos para cada establecimiento igual a la cantidad de observaciones que asistieran realmente al colegio en la base de datos utilizada para la estimación. En los casos que se predijo mayor demanda que capacidad, se priorizó de manera aleatoria para definir quienes quedaban dentro⁷. Se asignó al siguiente colegio, según preferencia, aquellas observaciones que quedaban sin cupo, repitiendo el proceso de manera iterativa hasta quedar todos asignados en algún establecimiento.

En vista del componente aleatorio de los ejercicios, ya sea por la definición de éstos, por las observaciones con “probabilidad de disponibilidad”, o por el proceso que regula la asignación para respetar los cupos, se calcularon 200 repeticiones con distintas semillas.

8.2 Resultados

Para evaluar el ajuste de la predicción de la segregación, se calculó el índice Duncan considerando como límites de vulnerabilidad los percentiles de ingreso del 10 al 90, con intervalos de 10. La Figura 8.1A representa la segregación real, mientras que la Figura

⁴La descripción de los 20 ejercicios realizados se desarrolla en el anexo I.

⁵Un modelo de elección discreta como el utilizado predice una probabilidad de elección para cada una de las opciones, por lo que no sería correcto simplemente hacer un ranking según probabilidad predicha. Se agradece el comentario de Juan Carlos Muñoz en este tema.

⁶Más detalles en anexo J.

⁷En la realidad, la sobre demanda es comúnmente solucionada con sistemas de selección. En este trabajo, la selección que realizan los colegios solo se considera en la modelación del proceso de elección (estimación de preferencias), mediante restricciones de disponibilidad, no en el proceso de distribución de cupos.

8.1B la segregación estimada según las elecciones predichas por el modelo. El color azul corresponde al grupo municipal, rojo subvencionado, verde privado, y amarillo todos los colegios⁸.

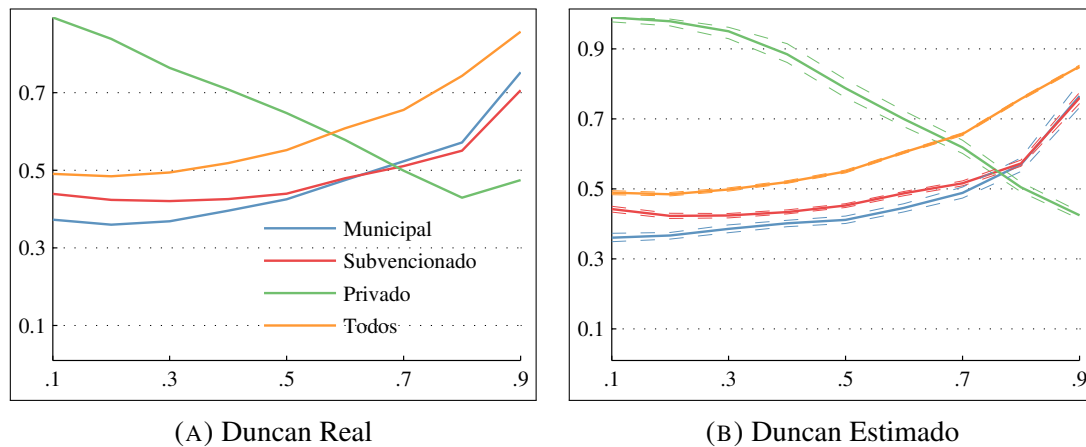


FIGURA 8.1. Índices de segregación Duncan (eje Y) para distintos percentiles de vulnerabilidad (eje X).

Se observa que los perfiles de segregación real (Figura 8.1A) son visualmente similares a los estimados (Figura 8.1B), lo que sugiere una buena capacidad del modelo estimado para predecir los niveles de estratificación. Se aprecia una tendencia a sobrestimar la segregación para el caso de colegios particulares.

Los resultados más importantes, referidos únicamente al sistema con financiamiento estatal, se resumen en la Tabla 8.1 y la Figura 8.2. La primera muestra los índices de segregación para los escenarios más relevantes, considerando la segregación del 30 % de menor ingreso. Además exhibe las disminuciones porcentuales en segregación, tomando en cuenta solo la porción de segregación que hay que explicar: la diferencia entre la segregación estimada y la segregación del escenario aleatorio⁹. Se exhiben los resultados de los

⁸La línea sólida de la figura 8.1B corresponde a la mediana de las 200 repeticiones calculadas, mientras que las punteadas a los cuantiles 0,1 y 0,9.

⁹Que coincide con la disminución porcentual en el “índice de segregación sistemática” propuesto en Carrington y Troske (1997). Winship (1977) argumenta que cuando se analizan las causas de la segregación, es pertinente considerar como línea base la segregación calculada en el escenario de asignación aleatoria.

diferentes índices de segregación propuestos. En la Figura 8.2, se muestra gráficamente la descomposición por incidencia de factores de forma independiente e interacción de éstos.

TABLA 8.1. Resumen resultados ejercicios según diferentes índices para variable dicotómica

	Índice			
	<i>D</i>	<i>H</i>	<i>R</i>	<i>S</i>
Cálculos de segregación				
Real	0.412	0.182	0.214	0.142
Residencial	0.317	0.109	0.126	0.088
Aleatoria	0.105	0.016	0.019	0.012
De las elecciones predichas por el modelo:				
Estimada	0.409	0.171	0.201	0.135
Sin copago	0.378	0.141	0.172	0.107
Sin segregación Urbana	0.385	0.151	0.178	0.119
Igual preferencias por currículum	0.332	0.124	0.149	0.096
Igual preferencias por SIMCE	0.379	0.152	0.178	0.120
Disminución porcentual				
Sin copago	10	19	16	23
Sin segregación urbana	8	13	13	13
Igual preferencias por currículum	25	31	29	32
Igual preferencias por SIMCE	10	12	12	12

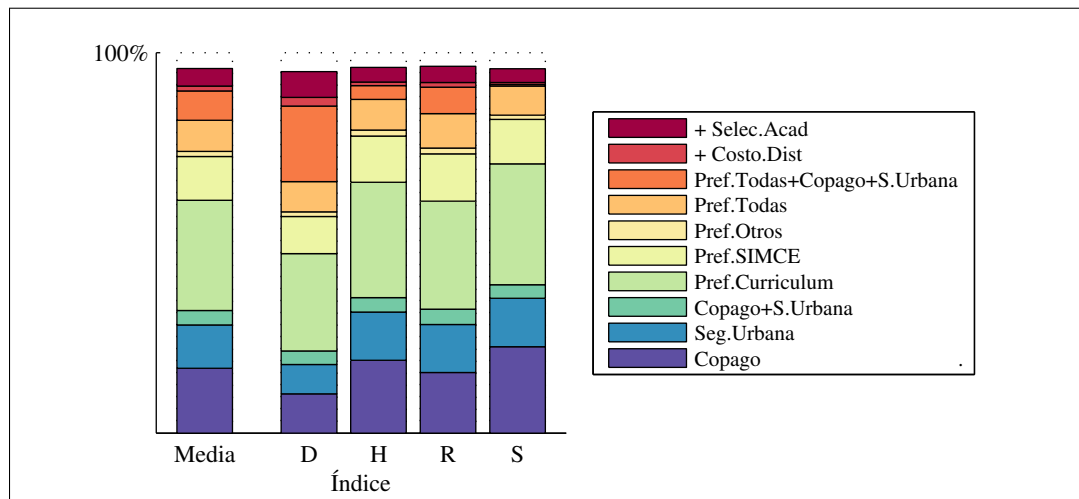


FIGURA 8.2. Descomposición de la segregación según diferentes índices

Dependiendo del indicador que se utilice, el copago explica entre un 10 % y 23 % de la segregación escolar, mientras que a la segregación residencial se le puede atribuir entre un 8 % y un 13 %. Se observa además que la segregación residencial es bastante menor a la escolar. Se aprecia que la eliminación de la selección académica no aporta en disminuir la segregación socioeconómica, de hecho la empeora entre un 2 % y un 3 %. Este resultado, como el de todos los otros ejercicios, se explica por las preferencias obtenidas en la modelación. Se estimó que los grupos medios tienen mayor preferencia por colegios con buen rendimiento y menor costo de viajar, por lo que una hipótesis es que utilizarían con mayor probabilidad los cupos de colegios de excelencia, desplazando a otros de menor ingreso, que con selección académica tenían prioridad por historial de buen rendimiento.

Se aprecia que la preferencia por colegios con formación técnico profesional tiene gran impacto en la segregación. Si todos tuvieran la misma valoración por tipo de enseñanza, se estima que la segregación del 30 % más vulnerable podría bajar entre un 22 % y un 27 %. Por último, se observa que la disparidad en valoración del SIMCE podría explicar entre un 10 % y un 12 % de la segregación¹⁰.

8.2.1 Índice de segregación sin grupo “vulnerable”

Dado que el problema de segregación no es exclusivo de los grupos de menores recursos, es prudente calcular un índice de segregación que no considere un grupo “vulnerable”, sino que intente explicar la segregación general. Siguiendo el trabajo Reardon (2011), se construyen los *rank-order segregation indices*¹¹. Los resultados se muestran en la tabla 8.2.

¹⁰Si se calcularan los mismos porcentajes para el ejercicio 15, en que las observaciones tienen igual valoración del SIMCE, distancia y tipo de enseñanza, la segregación se estima que bajaría entre un 34 % y un 40 %.

¹¹Para detalles de la construcción del indicador ver anexo D y sección 2.2.2 para mayor información sobre el índice. De los tres índices propuestos por el autor: *rank-order information theory index*, *rank-order variance ratio index* y *rank-order square root index*, se eligió el último por la similitud de los perfiles del *square root index* al los del índice Duncan, a pesar de que tienen diferente escala e interpretación.

Los resultados en cambios de segregación de los diferentes escenarios siguen la misma dirección que los de la sección anterior, aunque varían en magnitud¹². Según estos cálculos, la segregación urbana explica entre un 11 % y un 14 % de la segregación escolar en el sistema municipal y subvencionado, mientras que el copago entre un 15 % y 16 %. Recordar que la segregación socioeconómica que evalúa este indicador no es la de un grupo vulnerables, sino que la pareja distribución de observaciones con distintos niveles de ingreso en los colegios. Nuevamente el tipo de enseñanza que imparte el colegio explica una parte importante de la segregación (un 26 %), como también lo hace la distinta valoración del SIMCE que tienen las familias (en torno al 8 %).

TABLA 8.2. Resumen resultados ejercicios según diferentes índices para variable continua

	Índice		
	<i>R-O H</i>	<i>R-O R</i>	<i>R-O S</i>
Cálculos de segregación			
Real	0.218	0.212	0.222
Residencial	0.178	0.179	0.171
Aleatoria	0.029	0.019	0.044
De las elecciones predichas por el modelo:			
Estimada	0.212	0.203	0.216
Sin copago	0.184	0.174	0.190
Sin segregación Urbana	0.191	0.184	0.192
Igual preferencias por currículum	0.164	0.154	0.172
Igual preferencias por SIMCE	0.197	0.187	0.204
Disminución porcentual			
Sin copago	15	16	15
Sin segregación urbana	11	11	14
Igual preferencias por currículum	26	27	26
Igual preferencias por SIMCE	8	9	7

¹²No son comparables en términos de grado de segregación pues la interpretación de los índices usados son distintas.

9. CONCLUSIONES

La segregación del sistema escolar es un problema complejo, y aunque esté demás decir, de soluciones complejas. Hoy poco se sabe de la magnitud de las razones que explicarían la segregación. Hay efectos que se traslapan, como el copago y la selección, o la segregación urbana y las restricciones de movilidad, por lo que el detalle no es trivial. Con la intención de aportar al debate, en este trabajo se estimó el impacto de algunos atributos del sistema actual en los niveles de segregación¹.

Se estimaron los atributos más relevantes en la elección de colegio, en base a un modelo de elección de discreta con capacidad de interiorizar la variabilidad en preferencias, sin necesidad de imponer una forma funcional. Además, se considerando restricciones en disponibilidad y selección académica. Precio, distancia, SIMCE y tipo de enseñanza son valorados de forma bastante heterogénea dentro de la población. Familias con más recursos ponen mayor énfasis en el rendimiento, les es menos costoso viajar, en los extremos actúan como si valoraran de forma positiva el cobro y evitan colegios con formación técnico profesional. Los resultados están en la línea de trabajos anteriores (Chumacero et al., 2011; Gallego y Hernando, 2008b), e implican que, incluso si no existiera segregación urbana ni pago en los colegios, la segregación sería altamente probable.

Utilizando estas preferencias y ejercicios de equilibrio parcial, se calculó la segregación del sector con aporte estatal que hay que explicar: la diferencia entre la segregación estimada y un sistema integrado. Los resultados muestran que entre 8 % y un 13 % se debe a la segregación urbana, entre un 10 % y 23 % al copago, y el resto a las propias preferencias de las familias, restricciones no consideradas y fricciones del sistema. Se desprende que eliminar el financiamiento compartido va en dirección correcta, pero que para un mayor impacto se debe considerar acompañar de mayor información, apoyo en el proceso de

¹Enfocado en educación media del Gran Santiago, y utilizando datos previos a la implementación de la ley de subvención preferencial.

elección, fiscalización efectiva de la selección², y tener presente otros factores que podrían explicar el residuo entre la segregación que habría sin el copago y la integración.

Una idea en este sentido sería un proceso de postulación a colegios centralizado, que garantice transparencia, más información, mayor apoyo en la decisión, asegure respeto de la ley, prohíba selección, y termine con la reputación de colegios que son “solo para algunos”³. Trabajos recientes han generado propuestas en esta línea; como Hernando y Niklitschek (2013a), quienes son enfáticos en la importancia de la aleatoriedad para priorizar en caso de sobredemanda⁴; y Elacqua y Santos (2013), quienes se basan en el ejemplo de Holanda. Que sea un portal web como lo es para las universidades en Chile o el sistema colegios públicos en Inglaterra; oficinas de postulación bien distribuidas donde se entregue información personalizada según residencia y preferencias, como el *Family Application Center* en Charlotte, North Carolina (Hastings et al., 2007); u otra alternativa, quedará por discutir. Otras ciudades en el mundo han implementado sistemas relacionados con resultados positivos⁵, un piloto podría ser una alternativa interesante para partir en Chile⁶.

Decisiones de la demanda, y la heterogeneidad en valoración de atributos, son muy relevantes en la configuración de la segregación. Políticas orientadas únicamente en la oferta son insuficientes para mejorar el nivel de integración. La evidencia muestra que el sistema no está estratificado únicamente por las restricciones en los grupos más pobres, las

²Ya lo decía un ex ministro de educación de la década de los noventa: “El sistema necesita fortalecer la capacidad de los padres para que sean ellos los que informadamente escojan las escuelas para sus hijos y no sean éstas las que decidan a qué estudiante admitir y cuáles no.” (Arellano, 2000).

³Que además ayude a regular las disparidades de gasto en marketing, y que disminuya la posibilidad de situaciones insólitas como pago de los sostenedores a las madres por matricular a sus hijos (Raczynski et al., 2011)

⁴Existe literatura reciente de mecanismos de asignación (Abdulkadiroglu, Che, y Yasuda, 2012; Abdulkadiroglu y Sonmez, 2003; Pathak, 2011), con resultados positivos en términos de utilidad. Carrasco et al. (2014) también postula una idea de priorización por tómbola.

⁵Algunos ejemplos son Londres, Nueva York o Boston.

⁶Tal como sugieren Elacqua y Santos (2013).

preferencias de las familias inducen gran parte de la segregación, lo que es más evidente a mayor ingreso. Construir un sistema educacional más integrado requiere quitar restricciones, apoyar en la decisión a los más vulnerables, y conjuntamente pensar en la valoración por atributos de los grupos medios y altos, los más segregados⁷. Se trata de crear las condiciones⁸ para que las elecciones de las familias no resulten en la estratificación que se produce hoy.

⁷En ningún caso se plantean éstos como requisitos suficientes. Otros aspectos y propuestas, como incentivos monetarios a tener una matrícula más diversa (“subvención a la integración”), deben considerarse.

⁸Que probablemente sea un sistema con reconocimiento transversal de calidad, sin restricciones de acceso, que premie la diversidad; ¡que difícil es pensar en acabar la segregación sin una estrategia a muy largo plazo!

BIBLIOGRAFÍA

Abdulkadiroglu, A., Che, Y.-K., y Yasuda, Y. (2012). *Expanding “choice” in school choice* (Discussion Papers n.º 0809-09). New York, NY: Columbia University, Department of Economics.

Abdulkadiroglu, A., y Sonmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, 93(3), 729–747.

Allen, R. (2007). Allocating pupils to their nearest secondary school: The consequences for social and ability stratification. *Urban Studies*, 44(4), 751–770.

Allen, R., y Vignoles, A. (2007). What should an index of school segregation measure? *Oxford Review of Education*, 33(5), 643–668.

Amador, F. J., González, R. M., y Ortúzar, J. d. D. (2005). Preference heterogeneity and willingness to pay for travel time savings. *Transportation*, 32(6), 627–647.

Arellano, J. P. (2000). *Reforma educacional: prioridad que se consolida*. Santiago de Chile: Editorial Los Andes.

Åslund, O., y Skans, O. N. (2009). How to measure segregation conditional on the distribution of covariates. *Journal of Population Economics*, 22(4), 971–981.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3), 244–263.

Azmat, G., y Garcia Montalvo, J. (2012). *The role of awareness, information gathering and processing in school choice* (Barcelona GSE Working Paper Series n.º 639). Barcelona: Barcelona Graduate School of Economics.

Ball, S. J. (1993). Education markets, choice and social class: the market as a class strategy in the UK and the USA? *British Journal of Sociology of Education*, 14(1), 3–19.

Başar, G., y Bhat, C. (2004). A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B: Methodological*, 38(10), 889–904.

Belfield, C. R., y Levin, H. M. (2002). The effects of competition between schools on educational outcomes: A review for the United States. *Review of Educational research*, 72(2), 279–341.

Berndt, E. K., Hall, B. H., y Hall, R. E. (1974). Estimation and inference in nonlinear structural models. En *Annals of economic and social measurement* (Vol. 3, pp. 103–116). Cambridge, MA: NBER.

Berry, S., Levinsohn, J., y Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 63(4), 841–890.

Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1), 34–48.

Boxall, P. C., y Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics*, 23(4), 421–446.

Bravo, D., Del Pino, G., Donoso, G., Manzi, J., Martínez, M., y Pizarro, R. (2010, Septiembre). *Resultados de la vinculación prueba SIMCE 2006 y PSU admisión 2009* (Inf. Téc.). Santiago de Chile: Comité Técnico Asesor, Consejo de Rectores de las Universidades Chilenas.

Breffe, W. S., y Morey, E. R. (2000). Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of Atlantic salmon fishing. *Marine Resource Economics*, 15(1), 1–20.

Briebe, D. (2013). *¿Es la segregación escolar un problema de justicia?* (Documento de Trabajo). Santiago de Chile: Horizontal.

Burgess, S., Greaves, E., Vignoles, A., y Wilson, D. (2010). *What parents want: school preferences and school choice* (DoQSS Working Paper n.º 09-01). London: Department of Quantitative Social Science, University of London.

Burgess, S., McConnell, B., Propper, C., y Wilson, D. (2007). The impact of school choice on sorting by ability and socioeconomic factors in English secondary education. En L. Woessmann y P. E. Peterson (Eds.), *Schools and the Equal opportunity problem* (pp. 273–292). Cambridge, MA.: The MIT Press.

Cantillo, V., y Ortúzar, J. d. D. (2005). A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological*, 39(7), 641–657.

Carrasco, A., Bogolasky, F., Flores, C., Gutiérrez, G., y San Martín, E. (2014). *Selección de estudiantes y desigualdad educacional en Chile: ¿qué tan coactiva es la regulación que la prohíbe?* (Proyecto FONIDE n.º 711286). Santiago de Chile: Centro de Estudios en Políticas y Prácticas en Educación y Centro de Políticas Públicas – Pontificia Universidad Católica de Chile.

Carrington, W. J., y Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, 15(4), 402–409.

Chumacero, R. A., Gallegos, J., y Paredes, R. D. (2014). *Entry, competition and academic performance in a vouchers context* (Working Paper). Santiago de Chile: Pontificia Universidad Católica de Chile, Department of Industrial Engineering.

Chumacero, R. A., Gómez, D., y Paredes, R. D. (2011). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. *Economics of Education Review*, 30(5), 1103–1114.

Cortese, C. F., Falk, R. F., y Cohen, J. K. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review*, 630–637.

de Lapparent, M. (2010). Latent class and mixed logit models with endogenous choice set formation based on compensatory screening rules. En S. Hess y A. Daly (Eds.), *Choice modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference* (pp. 371–396). Bingley: Emerald Group Publishing.

Dempster, A. P., Laird, N. M., y Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.

Duncan, O., y Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210–217.

Elacqua, G. (2012). The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development*, 32(3), 444–453.

Elacqua, G., y Martínez, M. (2010). *When expectations are not fulfilled: School choice in Chile* (Documento de Trabajo). Santiago de Chile: Universidad Diego Portales, Instituto de Políticas Públicas.

Elacqua, G., y Santos, H. (2013). *Segregación y polarización en el sistema educacional chileno y recientes tendencias: ¿qué ha sucedido con los grupos medios?* (Documento de Referencia n.º 1). Santiago de Chile: Espacio Público.

Farley, R. (1977). Residential segregation in urbanized areas of the United States in 1970: An analysis of social class and racial differences. *Demography*, 14(4), 497–518.

Flores, C., y Carrasco, A. (2013). *(Des)igualdad de oportunidades en la escuela: preferencias, libertad de elección y segregación escolar* (Documento de Referencia n.º 2). Santiago de Chile: Espacio Público.

Friedman, M. (1962). *Capitalism and freedom*. Chicago: University of Chicago Press.

Gallego, F. A. (2006). *Voucher-school competition, incentives, and outcomes: Evidence from Chile* (Processed). Boston, MA: Massachusetts Institute of Technology.

Gallego, F. A., García, C., Lagos, F., y Stekel, Y. (2008). *Estudio longitudinal: evaluación de impacto del sistema nacional de evaluación de desempeño (SNED) en la educación subvencionada* (Inf. Téc.). Santiago de Chile: Pontificia Universidad Católica de Chile, Programa de Políticas Públicas.

Gallego, F. A., y Hernando, A. E. (2008a). On the determinants and implications of school choice: semi-structural simulations for Chile. *Economía*, 9(1), 197–239.

Gallego, F. A., y Hernando, A. E. (2008b). *School choice in Chile: Looking at the demand side* (Documento de Trabajo n.º 356). Santiago de Chile: Pontificia Universidad Católica de Chile, Instituto de Economía.

Gallego, F. A., Lagos, F., y Stekel, Y. (2012). *Types of information and school choice: An experimental study in the chilean voucher system [preliminary and incomplete]* (Inf. Téc.). Latin American and Caribbean Economic Association (LACEA) & Latin American Meeting of the Econometric Society (LAMES) 2012.

Gallego, F. A., y Sapelli, C. (2007). El financiamiento de la educación en Chile: una evaluación. *Revista de Pensamiento Educativo*, 40(1), 263–284.

García-Huidobro, J. E. (2007). Desigualdad educativa y segmentación del sistema escolar. consideraciones a partir del caso chileno. *Pensamiento Educativo*, 40(1), 65–86.

Glaeser, E. L., y Vigdor, J. L. (2003). Racial segregation: promising news. En B. Katz y L. R. E (Eds.), *Redefining urban and suburban America: Evidence from census 2000* (pp. 211–234). Washington, DC: Brookings Institution Press.

Gorard, S. (1999). Well. that about wraps it up for school choice research: a state of the art review. *School Leadership & Management*, 19(1), 25–47.

Gorard, S., y Taylor, C. (2002). What is segregation? A comparison of measures in terms of “strong” and “weak” compositional invariance. *Sociology*, 36(4), 875–895.

Grau, N. (2014, Abril). *A dynamic model of elementary school choice* (Manuscrito no publicado). University of Pennsylvania.

Greene, W. H. (2001). *Fixed and random effects in nonlinear models* (NYU Working Paper n.º EC-01-01). New York, NY: New York University Stern School of Business.

Greene, W. H., y Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), 681–698.

Greene, W. H., y Hensher, D. A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*, 37(3), 413–428.

Gupta, S., y Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 128–136.

Hastings, J. S., Van Weelden, R., y Weinstein, J. (2007). *Preferences, information, and parental choice behavior in public school choice* (NBER Working Paper n.º 12995). Cambridge, MA: National Bureau of Economic Research.

- Hastings, J. S., y Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4), 1373–1414.
- Hernando, A. E., y Niklitschek, K. (2013a). *Medidas para disminuir los niveles de segregación escolar* (Documento de Trabajo). Santiago de Chile: Horizontal.
- Hernando, A. E., y Niklitschek, K. (2013b). *Segregación escolar, financiamiento compartido y el sector educacional chileno*. (Documento de Trabajo). Santiago de Chile: Horizontal.
- Hess, S., Ben-Akiva, M., Gopinath, D., y Walker, J. L. (2011). Advantages of latent class over continuous mixture of logit models. *Unpublished*.
- Hess, S., y Polak, J. W. (2005). Accounting for random taste heterogeneity in airport choice modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 1915(1), 36–43.
- Heyneman, S. P. (2004). International education quality. *Economics of Education Review*, 23(4), 441–452.
- Hsieh, C.-T., y Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. *Journal of Public Economics*, 90(8–9), 1477–1503.
- Hutchens, R. (2001). Numerical measures of segregation: desirable properties and their implications. *Mathematical Social Sciences*, 42(1), 13–29.
- Hutchens, R. (2004). One measure of segregation. *International Economic Review*, 45(2), 555–578.
- James, D. R., y Taeuber, K. E. (1985). Measures of segregation. En N. B. Tuma (Ed.), *Sociological methodology* (pp. 1–31). San Francisco, CA: Jossey-Bass.

Jenkins, S. P., Micklewright, J., y Schnepf, S. V. (2008). Social segregation in secondary schools: how does England compare with other countries? *Oxford Review of Education*, 34(1), 21–37.

Kamakura, W. A., y Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4), 379–390.

Ladd, H. F. (2002). School vouchers: a critical view. *Journal of Economic Perspectives*, 16(4), 3–24.

Larrañaga, O., y Sanhueza, C. (2007). *Residential segregation effects on poor's opportunities in chile* (Documentos de Trabajo n.º 259). Santiago de Chile: Departamento de Economía, Universidad de Chile.

Larroulet, C. (2011). *Análisis de la movilidad escolar en chile* (Tesis de Master no publicada). Instituto de Economía, PUC.

Levin, H. M. (1998). Educational vouchers: effectiveness, choice, and costs. *Journal of Policy Analysis and Management*, 17(3), 373–392.

Libertad y Desarrollo. (2012). *Presupuesto para educación 2013: Fuerte aumento en Áreas clave* (Temas Públicos n.º 1083). Santiago de Chile: Autor.

Lindbom, A. (2010). School choice in Sweden: effects on student performance, school costs, and segregation. *Scandinavian Journal of Educational Research*, 54(6), 615–630.

Madero, C., y Madero, I. (2012). Elección escolar y selección estudiantil en el sistema escolar chileno ¿quién elige a quién?: el caso de la educación católica. *Revista mexicana de Investigación Educativa*, 17(55), 1267–1295.

Magidson, J., y Vermunt, J. K. (2007). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. En I. Sawtooth Software (Ed.), *Proceedings of the sawtooth software conference* (pp. 139–154). Santa Rosa, CA.

- Martínez, F., Aguila, F., y Hurtubia, R. (2009). The constrained multinomial logit: a semi-compensatory choice model. *Transportation Research Part B: Methodological*, 43(3), 365–377.
- Massey, D. S., y Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2), 281–315.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. En P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York, NY: Academic Press.
- McLachlan, G., y Krishnan, T. (2007). *The em algorithm and extensions* (Vol. 382). Hoboken, NJ: John Wiley & Sons.
- Mizala, A., González, P., y Romaguera, P. (2004). *Vouchers, inequalities and the Chilean experience* (Working Paper). Santiago de Chile: Universidad de Chile, Centro de Economía Aplicada (CEA).
- Mizala, A., y Torche, F. (2012). Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development*, 32(1), 132–144.
- Mizala, A., y Torche, F. (2013). *¿Logra la subvención escolar preferencial igualar los resultados educativos?* (Documento de Referencia n.º 9). Santiago de Chile: Espacio Público.
- Mora, R., y Ruiz-Castillo, J. (2003). *An evaluation of an entropy based index of segregation* (UC3M Working Paper). Madrid: Universidad Carlos III de Madrid. Departamento de Economía.
- Neal, D. (2002). How vouchers could change the market for education. *Journal of Economic Perspectives*, 16(4), 25–44.

Neilson, C. (2013). *Targeted vouchers, competition among schools, and the academic achievement of poor students* (Job Market Paper). New Haven, CT: Yale University.

Nerella, S., y Bhat, C. R. (2004). Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*, 1894(1), 11–19.

OECD. (2010). *Pisa 2009 results: Overcoming social background? equity in learning opportunities and outcomes (volume ii)*. OECD Publishing.

OECD. (2013a). *Education at a glance 2013: Oecd indicators*. OECD Publishing.

OECD. (2013b). *Pisa 2012 results: Excellence through equity: Giving every student the chance to succeed (volume ii)*. OECD Publishing.

Ortúzar, J. d. D., y Willumsen, L. G. (2011). *Modelling transport*. Chichester: John Wiley & Sons.

Östh, J., Andersson, E., y Malmberg, B. (2013). School choice and increasing performance difference: a counterfactual approach. *Urban Studies*, 50(2), 407–425.

Pagliara, F., y Timmermans, H. (2009). Choice set generation in spatial contexts: a review. *Transportation Letters*, 1(3), 181–196.

Paredes, R. D., Volante, P., Zubizarreta, J. R., y Opazo, M. (2013). Financiamiento compartido en la educación subvencionada chilena. En I. Irrázaval, M. d. I. Á. Morandé, y M. Letelier (Eds.), *Propuestas para Chile, concurso políticas públicas UC 2013* (pp. 51–83). Santiago de Chile: Pontificia Universidad Católica de Chile.

Pathak, P. A. (2011). The mechanism design approach to student assignment. *Annual Review of Economics*, 3(1), 513–536.

Provencher, B., y Bishop, R. C. (2004). Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. *Journal of Environmental Economics and Management*, 48(1), 793–810.

Raczynski, D., Hernández, M., Kegevic, L., Roco, R., González, R., y Peña, J. (2011). *El paso de la enseñanza básica a la media en estratos bajos: un reto a la igualdad de oportunidades educativas* (Proyecto FONIDE n.º F511066). Santiago de Chile: Asesorías para el Desarrollo.

Raczynski, D., Salinas, D., De la Fuente, L., Hernández, M., y Lattz, M. (2010). *Hacia una estrategia de validación de la educación pública-municipal: imaginarios, valoraciones y demandas de las familias* (Proyecto FONIDE n.º F310827). Santiago de Chile: Asesorías para el Desarrollo.

Reardon, S. F. (2009). Measures of ordinal segregation. *Research on Economic Inequality*, 17, 129–155.

Reardon, S. F. (2011). *Measures of income segregation* (Working Paper). Stanford, CA: Stanford Center for Education Policy Analysis.

Reardon, S. F., y Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32(1), 33–67.

Rojas, M. T. (2009). ¿Es posible imaginar escuelas más integradas socialmente? Respuestas de directivos, docentes y familias frente a una normativa que promueve la integración social. *Docencia*, 39, 31–29.

Sapelli, C., y Torche, A. (2002). Subsidios al alumno o a la escuela: efectos sobre la elección de colegios públicos. *Cuadernos de Economía*, 39(117), 175–202.

Schneider, M., Elacqua, G., y Buckley, J. (2006). School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management*, 25(3), 577–601.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3–55.

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*, 41(22), 2915–2924.

Swait, J., y Ben-Akiva, M. (1986). Analysis of the effects of captivity on travel time and cost elasticities. En *Behavioural research for transport policy: The 1985 international conference on travel behaviour* (pp. 119–134). Noordwijk: VNU Science Press BV.

Theil, H., y Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology*, 1(2), 187–193.

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Treviño, E., Salazar, F., y Donoso, F. (2011). ¿Segregar o incluir?: Esa no debería ser una pregunta en educación. *Docencia*, 45.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, 79(4), 281.

Valenzuela, J. P., y Allende, C. (2012). *Logros en liceos públicos de excelencia en Chile: ¿Valor agregado o sólo descreme de la elite?* (Documento de trabajo). Santiago de Chile: Universidad de Chile, Centro de Investigación Avanzada en Educación.

Valenzuela, J. P., Bellei, C., y de los Ríos, D. (2008). *Evolución de la segregación socioeconómica de los estudiantes chilenos y su relación con el financiamiento compartido* (Proyecto FONIDE n.º 211). Santiago de Chile: Universidad de Chile, Programa de Investigación en Educación.

Valenzuela, J. P., Bellei, C., y de los Ríos, D. (2010). Segregación escolar en Chile. En S. Martinic y G. Elacqua (Eds.), *¿fin de ciclo? cambios en la gobernanza del sistema educativo* (pp. 209–229). Santiago de Chile: UNESCO – Pontificia Universidad Católica de Chile.

Valenzuela, J. P., Villalobos, C., y Gómez, G. (2013). *Segregación y polarización en el sistema escolar chileno y recientes tendencias: ¿qué ha sucedido con los grupos medios?* (Documento de Referencia n.º 3). Santiago de Chile: Espacio Público.

Walker, J. L., y Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences*, 43(3), 303–343.

Walker, J. L., y Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1), 77–101.

Wen, C.-H., y Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review*, 46(2), 211–221.

Williams, H., y Ortúzar, J. d. D. (1982). Behavioural theories of dispersion and the misspecification of travel demand models. *Transportation Research Part B: Methodological*, 16(3), 167–219.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *The Teachers College Record*, 112(4), 1008–1037.

Winship, C. (1977). A revaluation of indexes of residential segregation. *Social Forces*, 55(4), 1058–1066.

Wu, C. (1983). On the convergence properties of the em algorithm. *The Annals of Statistics*, 11(1), 95–103.

Zoloth, B. S. (1976). Alternative measures of school segregation. *Land Economics*, 278–298.

ANEXOS

A. BASE DE DATOS

La base de datos fue construida a partir de datos del SIMCE 2006 y la PSU 2009. Se consideraron 37.859 (798 colegios) de un total de 85.284 observaciones (890 colegios) de la base del Cuestionario para Padres SIMCE 2006 de la RM , por las siguientes razones:

1. No tener ubicación geográfica por rut omitido, rut incorrecto, no inscrito en PSU, imposibilidad de georeferenciar su dirección u otro (34.433 obs.)¹.
2. Asistir a colegios fuera del área geográfica definida para la oferta de colegios (3.504 obs.).
3. Tener un valor omitido en variable relevante para estimación (8.956 obs.).
4. Vivir fuera de los límites establecidos para ser considerado como que sus alternativas disponibles son en su mayoría del Gran Santiago (532 obs.).

A.1 Área geográfica definida para el trabajo

El área geográfica definida para la oferta de colegios del trabajo comprende la Provincia de Santiago² más las comunas Puente Alto, San Bernardo, Padre Hurtado y Pirque. El criterio fue dejar fuera aquellas comunas que parecieran aisladas³ del resto en términos de oferta educacional. En la Figura A.1 se aprecian en azul los colegios dentro del área definida y en rojo los fuera del área. Así mismo, se limitó la ubicación de los hogares a las latitudes $-33,3^{\circ}$ y $-33,7^{\circ}$, y longitud mayor a $-70,84^{\circ}$, lo que se observa en la Figura A.2 (azul corresponde a los hogares que están dentro de los límites, rojo los que están

¹Con la información disponible no se puede saber la cuánto se debe a cada factor. 34.433 son las observaciones que no se puede hacer matching usando el rut entre la base SIMCE y las georeferencias hechas a partir de la base PSU.

²Incluye las comunas Santiago, Cerrillos, Cerro Navia, Conchalí, El Bosque, Estación Central, Huechuraba, Independencia, La Cisterna, La Florida, La Granja, La Pintana, La Reina, Las Condes, Lo Barnechea, Lo Espejo, Lo Prado, Macul, Maipú, Ñuñoa, Pedro Aguirre Cerda, Peñalolén, Providencia, Pudahuel, Quilicura, Quinta Normal, Recoleta, Renca, San Joaquín, San Miguel, San Ramón y Vitacura.

³No se utilizó ningún criterio formal, solo inspección visual.

fuera). El área de los individuos es más extensa que la definida para oferta educacional, y se estimó conveniente fijarla pues se eliminó la oferta de colegios en las periferias de Santiago, los que podrían haber sido alternativas relevantes para hogares ubicados a las afueras de Santiago.

FIGURA A.1. Oferta de colegios enseñanza media en la Región Metropolitana

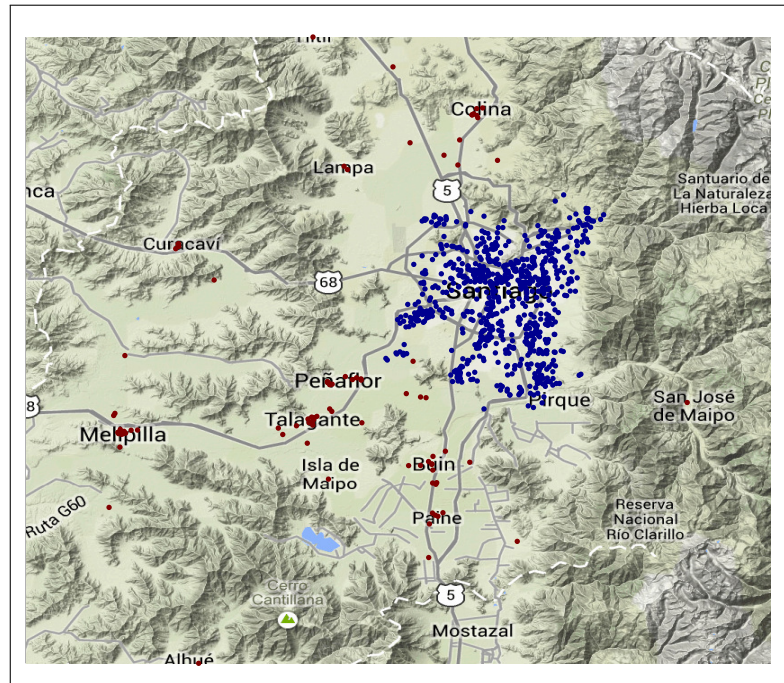
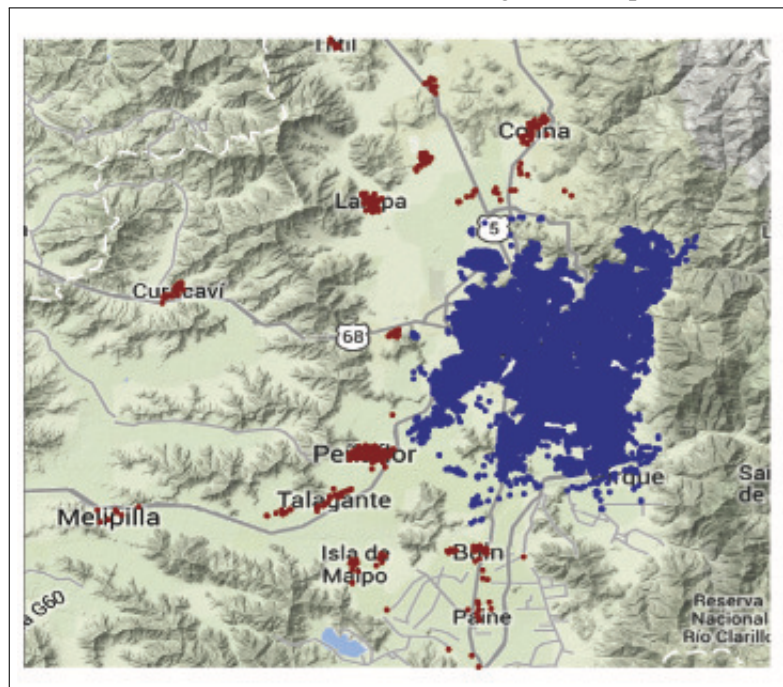


FIGURA A.2. Observaciones en Región Metropolitana



B. INGRESOS IMPUTADOS

Con el fin de poder estimar parámetros y efectos marginales, dependientes del nivel de ingreso con el menor sesgo posible, se procedió a imputar el ingreso para las declaraciones del Cuestionario de Padres SIMCE 2006 que estuvieran en la categoría más alta (opción “sobre \$1.800.000), y por ende truncada.

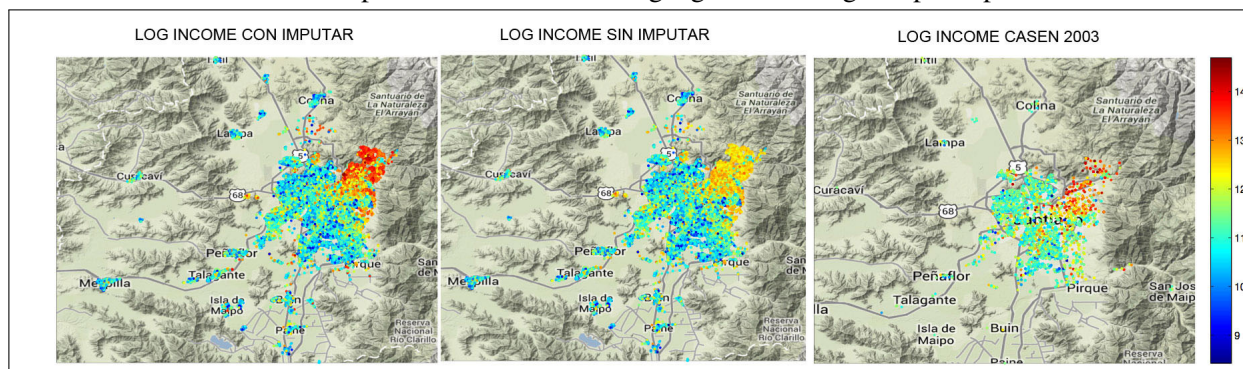
La estrategia utilizada fue promediar el ingreso per cápita de los vecinos de residencia de la observación, según la encuesta Casen 2003¹, con el ingreso per cápita promedio de las familias asistiendo al mismo colegio, con padres de igual nivel educacional². Estos últimos datos se obtuvieron de las encuestas Casen 2003, 2006 y 2009. De un total de 3.694 observaciones con declaración de ingreso en la cota superior se corrigieron 2.910 valores, 759 no se reemplazaron porque el valor calculado era menor al declarado (\$1.800.000 dividido por el número de personas en el hogar) y 25 porque no se pudo construir ingreso debido a falta de datos.

En la Figura B.1 se observa la diferencia de considerar o no los ingresos imputados, y la comparación con la distribución geográfica del ingreso según la CASEN 2003. Es visualmente pertinente la imputación, pues de esta manera se incluyen algunos valores de la parte alta de la distribución que antes estaban truncados. Cabe notar que los supuestos en que se basa la estrategia de homogeneidad de ingresos en vecindario y colegios no son del todo reales, pero dentro de las alternativas evaluadas parecía ser la más adecuada.

¹Esto fue posible gracias a que se tenía la georeferencia de las observaciones, trabajo realizado en Chumacero et al. (2011). Se consideró un promedio ponderado por una función decreciente en la distancia $\left(\left(\frac{1}{dist+60}\right)^3\right)$.

²El trabajo de búsqueda de educación de los padres no es trivial en el contexto de la Casen, pues no es un dato que se declare a nivel de observación, si no que hay que buscar al apoderado en la base

FIGURA B.1. Comparación de distribución geográfica del ingreso per cápita



C. ANÁLISIS COMUNAL

La oferta de colegios en educación media es bastante variada según comuna, tanto en cantidad como en tipo y calidad. La Figura C.1 muestra Santiago con un círculo en cada comuna, donde el color representa el SIMCE promedio de todos los colegios y el radio la cantidad de colegios (lo que no es lo mismo que la cantidad de cupos o matrícula). Se observa que las comunas con mejores colegios son las del sector nor-oriental, y las con peor rendimiento las del sur y nor-occidental.

Se procedió a realizar un análisis desagregado a nivel de comuna para conocer sobre la oferta comunal y los desplazamientos por motivo de estudio. Dado que las observaciones de la base no contenían identificador del municipio pero sí georeferenciación, se estimó la comuna de residencia de cada individuo a partir de la del colegio más cercano¹. Se calculan los residentes (o habitantes) por comuna como el número de observaciones de la base que residan en la comuna, similar para los estudiantes por comuna. El ingreso promedio de la comuna es la media de los ingresos de los residentes. El SIMCE e ingreso promedio por tipo de colegio son medias ponderadas según tamaño, del SIMCE e ingreso promedio de los apoderados de cada colegio respectivamente.

De la Figura C.2 y la Tabla C.1, se observa que las comunas que retienen en sus colegios a un mayor porcentaje de sus residentes son Santiago, San Bernardo, Puente Alto y Maipú, con al menos dos tercios (criterio de orden de la tabla). Mientras que las que no logran ofrecer alternativas atractivas ni para el 20 % son Lo Prado, Cerro Navia, Lo Espejo, Pedro Aguirre Cerda y La Granja. A excepción de Santiago, los municipios con mejor retención son comunas grandes y periféricas, lo que al menos geográficamente hace más probable este resultado. Peñalolén y Pudahuel cumplen con características similares, pero sus estudiantes tienden a salir bastante más de la comuna. El caso de Santiago se

¹ Si bien esta estrategia puede no ser muy efectiva en las fronteras, mediante inspección visual el resultado pareciera bastante bueno.

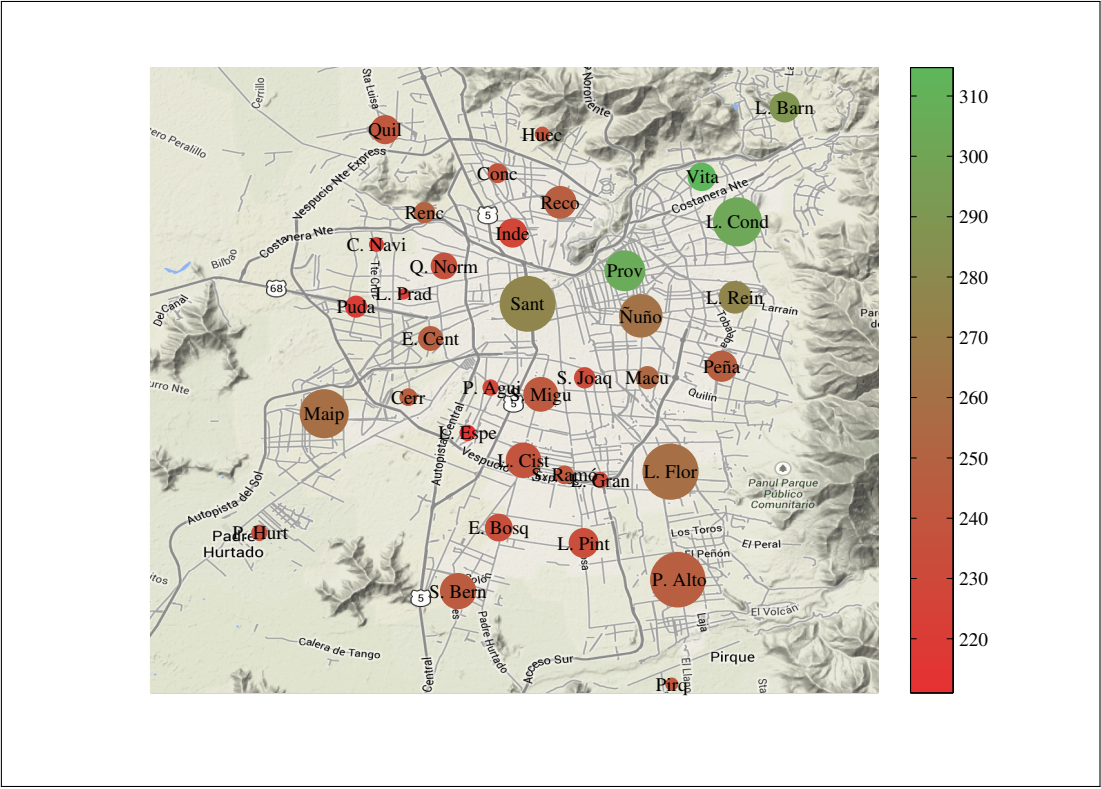


FIGURA C.1. Cantidad de colegios y SIMCE promedio por comuna

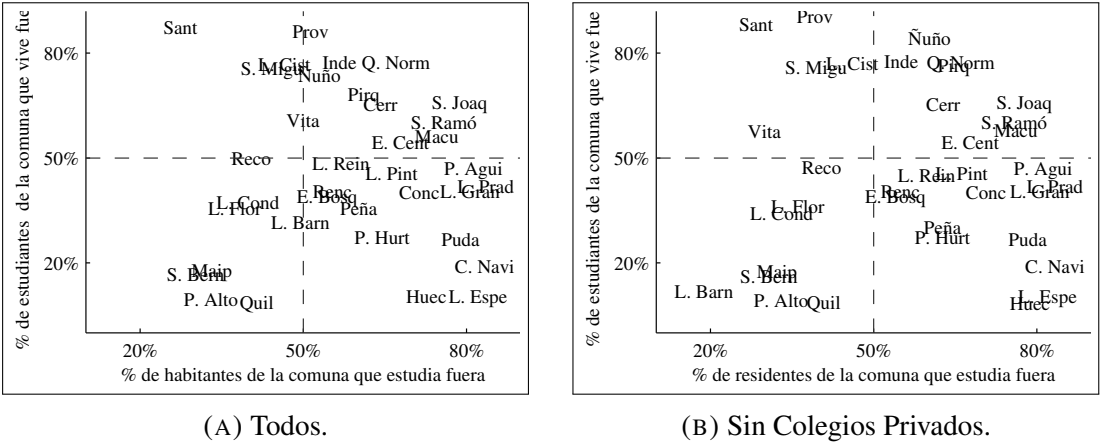


FIGURA C.2. Relación entre cantidad de alumnos que emigran e inmigran por comuna.

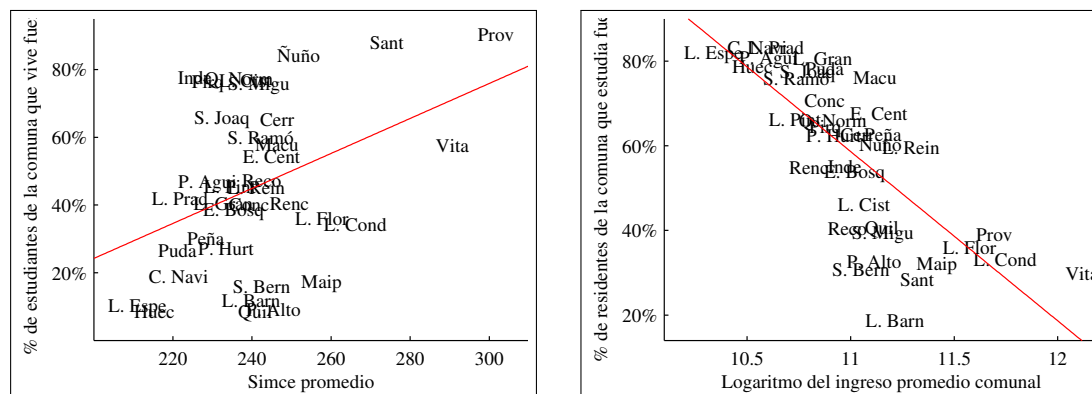
puede explicar por la extensa oferta educativa, tiene el segundo lugar después de la Florida en cantidad de colegios, con bastante heterogeneidad de opciones.

Las comunas que reciben mayor cantidad de estudiantes de otros municipios son Santiago, Providencia, La Cisterna, Independencia, Quinta Normal y San Miguel, donde más de tres cuartos de la matrícula corresponde a alumnos no residentes. Escenario opuesto se observa en Puente Alto, Quilicura, Huechuraba y Lo Espejo, donde esta cifra es menor a un 10 %. Nuevamente parte de la diferencia se podría explicar por acceso, las primeras son comunas centrales, mientras que las segundas todas periféricas. Casos como el de Estación Central, Macul y Recoleta; que siendo comunas más o menos centrales los alumnos de fuera representan menos del 60 %, indican que existen otros factores adicionales a la geografía.

En base a la Figura C.2B, que omite las observaciones de colegios privados, se podría clasificar en forma burda al agregado de colegios subvencionados y municipales de una comuna como: atractivos para todos (cuadrante II): Santiago, Providencia, San Miguel, Vitacura; atractivos solo para los residentes (cuadrante III): Lo Barnechea, Puente Alto, San Bernardo; atractivos solo para los de fuera (cuadrante I): Quinta Normal, San Joaquín; y los no atractivos (cuadrante IV): Huechuraba, Lo Espejo, Cerro Navia y Pudahuel.

Sin incluir las observaciones de colegios privados, la correlación entre el porcentaje de estudiantes de la comuna que reside fuera y el SIMCE promedio se observa en la Figura C.3A. El comportamiento general es el esperado, comunas con mayor SIMCE tienen proporciones más altas de estudiantes de otros municipios, pero existen bastantes casos que se alejan de esta tendencia. En Independencia, Quinta Normal y San Miguel, con SIMCE menor a la media, el porcentaje de alumnos afuerinos es sobre un 70 %, mientras que en La Florida y Las Condes, con SIMCEs mejores que el promedio, la proporción en cuestión es cercana a un tercio.

Una relación bastante más estricta se obtiene del porcentaje de estudiantes de estudio en una comuna distinta a la de residencia comparando con el logaritmo del ingreso comunal. Y es simple: comunas más ricas retienen mucho mejor a sus residentes escolares que municipios más pobres.



(A) % estudiantes de fuera vs SIMCE promedio(B) % Residentes salen vs ingreso promedio comuna

FIGURA C.3. Correlaciones de variables comunales (no incluye observaciones de colegios privados)

No está demás destacar que es necesario un análisis multivariado, hay bastante elementos que no se están considerando en los análisis anteriores, se deja para futuros desarrollos. Lo que se presenta es un primer paso para explorar la movilidad entre comunas, aspecto importante para entender la segregación.

C.1 Segregación

En una extensión de la sección 4, se calcularon los índices descompuestos por comuna. Si en la Tabla C.1 se presentaron indicadores sobre la oferta comunal, acá el interés está centrado en qué tan segregada está esa oferta.

La Tabla C.2 contiene los cálculos de segregación escolar y urbana, utilizando el índice de teoría de la información. Las columnas correspondientes a la segregación urbana están subdivididas en cuatro grupos. El primero representa el ejercicio de asignar las observaciones a su colegio más cercano (Allen, 2007); y los otros tres son el resultado de dividir la ciudad en cuadrados de distintos tamaños, donde cada uno representa una unidad de análisis (se podría interpretar cada uno como un colegio, ver Figura C.4). Para cada

TABLA C.1. Datos de matrícula y colegios a nivel comunal.

Comuna				Municipal			Subvencionado			Privado			Todos				
	Cantidad de Residentes	Cantidad de Estudiantes	Ingreso Promedio (M\$)	Cantidad de Colegios	Promedio SIMCE	Ingreso Promedio (M\$)	Cantidad de Colegios	Promedio SIMCE	Ingreso Promedio (M\$)	Cantidad de Colegios	Promedio SIMCE	Ingreso Promedio (M\$)	Cantidad de Colegios	Promedio SIMCE	Ingreso Promedio (M\$)	% Estudia Afuera	% Vive Afuera
Santiago	960	5511	130	22	276	78	37	271	91	10	317	269	69	277	95	27	87
San Bernardo	1620	1357	79	8	228	42	20	262	91	1	312	304	29	245	72	30	17
Puente Alto	4196	3124	80	7	226	47	55	250	72	6	298	301	68	246	73	33	10
Maipú	4079	3314	97	7	249	50	42	259	98	4	314	274	53	259	95	33	18
La Florida	2911	2833	126	7	236	53	57	263	118	6	300	279	70	259	112	37	36
Las Condes	2169	2082	490	5	262	114	9	268	131	39	317	613	53	303	480	40	37
Recoleta	885	1053	79	10	230	48	12	264	77	2	310	318	24	246	74	40	50
Quilicura	1178	756	85	2	218	37	16	247	78	1	322	556	19	242	76	41	9
San Miguel	606	1389	130	6	248	56	19	239	76	2	285	364	27	242	75	44	76
La Cisterna	602	1393	96	4	252	49	23	236	67	1	306	348	28	239	65	47	77
Lo Barnechea	877	648	663	3	223	55	4	253	90	14	320	694	21	288	445	49	32
Vitacura	690	881	635	2	289	164	3	294	214	13	322	699	18	315	573	50	61
Providencia	647	2275	345	5	318	111	8	262	141	24	315	542	37	306	260	51	86
Nuñoa	1019	1810	239	12	258	68	8	241	72	22	306	337	42	263	123	53	74
El Bosque	1328	991	65	3	212	37	14	242	68	-	-	-	17	235	61	54	39
Renca	699	524	61	1	260	47	9	247	49	-	-	-	10	249	49	55	40
Independencia	413	792	92	4	241	65	14	224	57	1	312	263	19	227	61	57	77
La Reina	1021	852	335	3	220	50	4	269	119	17	311	413	24	277	252	57	48
Peñalolén	1344	830	182	5	210	35	9	236	85	8	306	425	22	248	159	60	36
Pirque	54	66	192	-	-	-	2	229	53	2	299	412	4	238	101	61	68
Cerrillos	349	359	87	1	216	37	6	250	65	-	-	-	7	246	62	64	65
Padre Hurtado	283	139	90	1	200	34	5	242	60	-	-	-	6	233	55	64	27
La Pintana	889	552	48	2	208	34	18	241	49	-	-	-	20	234	46	66	46
Quinta Normal	638	929	81	7	232	46	9	246	73	-	-	-	16	237	55	67	77
Estación Central	996	703	98	3	223	44	10	259	85	1	295	335	14	247	82	68	54
Conchalí	736	352	71	4	239	51	5	239	54	-	-	-	9	239	53	71	40
Huechuraba	504	154	137	1	205	21	2	219	43	2	316	424	5	238	126	73	10
Macul	624	362	125	2	212	36	8	268	87	2	303	284	12	254	98	75	56
San Ramón	451	273	56	4	213	37	4	271	55	-	-	-	8	242	46	76	60
San Joaquín	485	302	83	2	214	33	8	238	54	-	-	-	10	232	49	79	66
Pudahuel	1547	444	85	2	211	37	9	224	58	-	-	-	11	221	53	79	27
La Granja	600	195	61	1	206	34	5	263	74	-	-	-	6	233	52	81	41
Pedro Aguirre Cerda	511	183	62	4	215	32	2	259	57	-	-	-	6	228	40	81	47
Lo Espejo	532	106	50	3	201	26	3	234	42	-	-	-	6	211	31	82	10
Cerro Navia	978	201	52	2	220	34	3	223	43	-	-	-	5	221	38	83	19
Lo Prado	438	124	62	1	211	38	2	268	57	-	-	-	3	222	42	84	42
Promedio	1052	1052	151	4	231	51	13	250	78	8	309	403	22	249	117	59	46

% Estudia Afuera: porcentaje de los residentes de la comuna que estudian en otra.

% Vive Afuera: porcentaje de los estudiante de la comuna que viven en otra.

El promedio no es ponderado, % de la muestra de Estudia Afuera es 51 %.

escenario se calculó el porcentaje de “vulnerables”, la segregación, y el aporte a la segregación *within*²; todo a nivel de comuna. A esto se agrega la segregación *within*, *between* y total a nivel de ciudad (últimas filas Tabla C.2).

²En la sección 4.1 se explica en que consiste la descomposición de un índice de segregación y conceptos de *within* y *between*

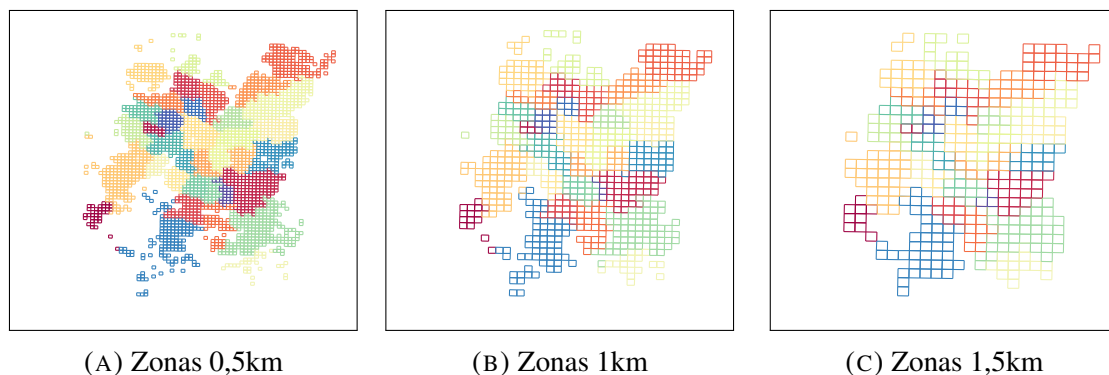


FIGURA C.4. División del Gran Santiago en cuadrados

Las comunas con sistemas escolares más segregados son Lo Barnechea y Huechuraba, mientras que las con menor nivel Conchalí y Lo Prado. Las que más aportan a la segregación escolar total son Santiago y La Florida (criterio de orden de la Tabla). Un 63 % de la segregación total se explica por lo que ocurre dentro de las comunas y un 37 % por cómo se reparten entre las comunas.

Respecto a la segregación residencial, se observa que, independiente de la unidad de análisis, las comunas más segregadas son Lo Barnechea y Las Condes. Las que influyen en mayor medida a la segregación de la ciudad son Puente Alto, La Florida, Maipú y Peñalolén. Se observa que la magnitud calculada para la segregación residencial depende fuertemente de lo que se utilice como unidad de análisis: mientras más pequeña la unidad, más alto el índice. Una extensión se observa en la Figura C.5, donde se muestra claramente la sensibilidad de los índices de segregación utilizados frente al tamaño de la unidad de análisis considerado. Cabe notar que, por construcción, la segregación *between* no depende del tamaño de la unidad de análisis (pues se calcula utilizando la comuna como unidad). En todos los escenarios la segregación educacional es más alta que la geográfica, y lo más interesante es que ambas presentan niveles similares de segregación *between*, pero difieren bastante en segregación *within*. Este último resultado implica que el sistema educacional probablemente tiene mecanismos de segregación más fuertes que el mercado residencial.

TABLA C.2. Descomposición de segregación escolar y urbana a nivel comunal.

Comuna	Escolar						Urbana								
				Colegio más cercano**			Zonas 1km			Zonas 0,5km			Zonas 1,5km		
	Vulnerables [%]	Segregación*	Aporte a seg. <i>within</i> [%]	Vulnerables [%]	Segregación*	Aporte a seg. <i>within</i> [%]	Vulnerables [%]	Segregación*	Aporte a seg. <i>within</i> [%]	Vulnerables [%]	Segregación*	Aporte a seg. <i>within</i> [%]	Vulnerables [%]	Segregación*	Aporte a seg. <i>within</i> [%]
Santiago	26	133	10.9	21	184	5.5	21	77	2.3	22	191	3.2	22	42	1.8
La Florida	20	236	8.6	23	97	9.3	24	106	9.0	23	164	8.6	22	58	7.3
Puente Alto	38	158	8.4	34	92	15.1	34	90	13.7	34	142	12.5	35	71	14.1
Maipú	24	146	6.9	25	56	7.8	26	72	9.6	25	113	8.5	25	53	9.6
Nuñoa	25	261	6.9	9	199	3.8	8	134	2.0	9	289	3.0	8	121	3.0
Peñalolén	27	475	6.0	28	127	6.2	27	184	8.0	28	270	7.0	26	122	6.4
San Bernardo	43	207	5.0	43	84	5.6	43	105	6.7	43	172	6.2	42	66	6.2
Las Condes	4	495	4.5	3	279	5.5	3	316	5.6	4	423	5.2	5	264	9.0
La Cisterna	51	171	4.3	33	49	1.1	35	41	0.9	33	90	1.1	32	11	0.3
Providencia	11	209	4.3	5	186	1.4	4	107	0.6	5	260	1.1	5	29	0.3
El Bosque	43	231	4.0	48	50	2.8	47	61	2.6	48	91	2.8	50	33	2.2
Recoleta	42	209	3.9	41	57	2.1	41	57	1.8	41	114	2.2	43	27	1.5
San Miguel	42	149	3.7	24	107	2.2	24	87	1.5	23	143	1.6	30	63	2.1
Lo Barnechea	12	517	3.1	9	380	6.3	9	340	5.2	8	430	3.4	6	381	5.7
La Reina	8	475	3.0	10	180	3.5	11	186	3.6	9	309	2.9	11	132	3.8
Estación Central	31	206	2.3	32	79	3.0	30	78	2.7	31	166	3.6	32	55	2.5
Macul	25	315	1.7	24	34	0.7	24	62	1.2	24	129	1.5	26	41	1.4
Quilicura	34	134	1.7	32	84	3.8	32	95	4.0	32	142	3.5	32	49	2.9
Huechuraba	45	524	1.4	41	185	3.8	41	246	4.9	41	303	3.2	41	288	6.2
Quinta Normal	47	81	1.4	34	29	0.7	33	21	0.5	33	60	0.8	34	26	0.8
Independencia	43	87	1.2	28	46	0.7	29	45	0.6	29	111	0.9	28	16	0.3
Pudahuel	49	123	1.0	31	53	3.1	31	62	3.2	31	120	3.7	30	57	4.6
Renca	52	91	0.9	46	5	0.1	47	12	0.3	45	53	0.8	48	13	0.5
Cerrillos	43	126	0.8	38	70	1.0	36	103	1.2	38	158	1.2	39	57	1.2
Vitacura	2	350	0.8	1	256	0.5	1	182	0.4	1	403	0.5	1	130	0.5
La Pintana	60	75	0.7	59	16	0.6	56	39	1.7	58	71	1.3	60	25	0.8
San Joaquín	53	107	0.6	38	30	0.6	35	61	1.2	38	159	1.6	36	17	0.3
La Granja	49	122	0.4	46	16	0.4	48	25	0.3	47	57	0.8	49	5	0.2
San Ramón	50	87	0.4	51	10	0.2	51	28	0.6	49	64	0.8	50	13	0.6
Pedro Aguirre Cerda	57	91	0.3	48	11	0.2	48	7	0.2	49	83	0.9	47	3	0.1
Padre Hurtado	60	140	0.3	38	106	1.2	43	124	1.2	35	279	1.9	39	102	1.5
Pirque	36	202	0.2	20	1	0.0	28	485	1.1	20	642	0.6	32	230	1.3
Cerro Navia	68	45	0.1	54	6	0.2	53	14	0.5	55	46	1.0	51	8	0.5
Conchalí	52	12	0.1	43	14	0.4	44	20	0.6	43	58	1.0	42	12	0.4
Lo Espejo	79	59	0.1	58	5	0.1	57	16	0.3	57	53	0.7	58	5	0.2
Lo Prado	65	19	0.0	44	16	0.3	47	6	0.1	45	27	0.3	47	0	0.0
Unidades	798			798			790			2520			403		
Segregación <i>within</i> *	167 (63 %)			71 (44 %)			78 (46 %)			131 (59 %)			54 (37 %)		
Segregación <i>between</i> *	96 (37 %)			92 (56 %)			90 (54 %)			91 (41 %)			92 (63 %)		
Segregación total*	263			162			168			222			146		

“Zonas Xkm” corresponde a la división de la ciudad en “manzanas” (gigantes) de lado de X kilómetros (ver Figura C.4).

* Índices de segregación multiplicados por 1000.

** Sin cupo máximo por colegio.

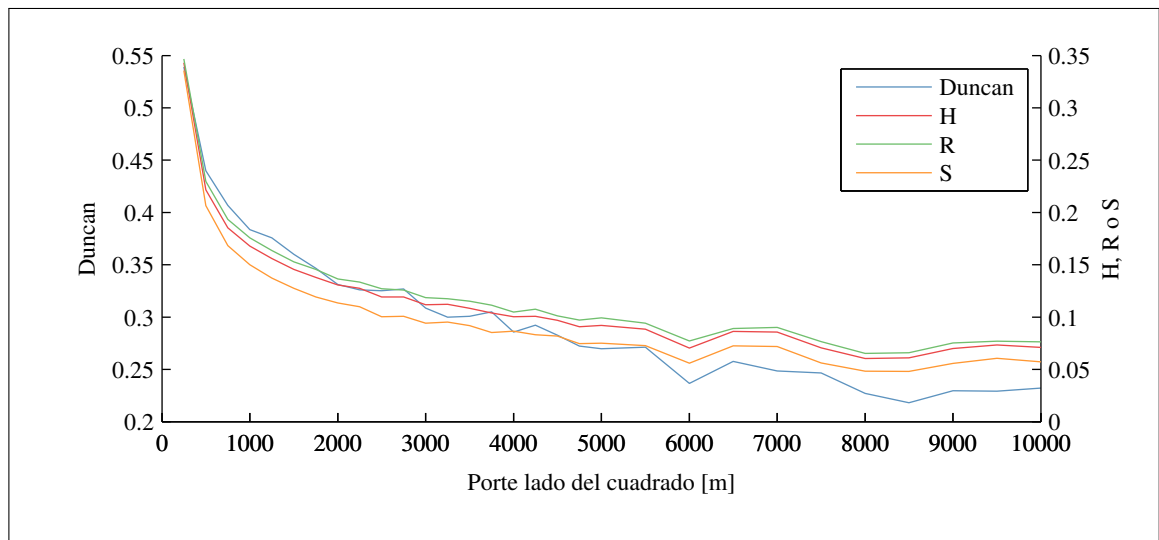


FIGURA C.5. Medición de segregación urbana (eje Y) según tamaño de la unidad (eje X)

D. ÍNDICE DE SEGREGACIÓN PARA VARIABLE CONTINUA

Reardon (2011) propone una serie de índices que permiten el cálculo de la segregación considerando una variable continua.

D.1 Propiedades

El autor menciona seis propiedades importantes que cumplen los índices propuestos:

Scale Interpretability:: Toma su valor máximo si en cada unidad no hay varianza en la variable, y mínimo si la distribución de la variable “imita” la distribución de la población.

Organizational Equivalence:: Si dos unidades con igual distribución en la variable de interés se juntan, entonces el índice no se ve afectado.

Population Size Invariance:: Si se multiplica la población por una constante positiva, entonces no hay efecto en la medición de segregación.

Rank Preserving Scale Invariance:: El indicador no se modifica frente a cambios en la variable de interés derivados de una multiplicación por algún número positivo o por adición de una constante igual para todas las observaciones.

Exchanges:: Un intercambio de dos observaciones entre dos unidades que produzca que las distribuciones de ambas unidades sean “más parecidas” disminuye la medición de segregación.

Ordered Exchanges:: En un intercambio como el anterior, mientras mayor sea la diferencia en ránking de las observaciones involucradas (es decir, mientras pertenezcan a percentiles más distantes), mayor será el impacto sobre el indicador.

D.2 Construcción

Como se menciona en la sección 2.2.2, el índice propuesto en Reardon (2011) se puede construir como un promedio ponderado de índices de segregación calculados en

base a distintos límites de vulnerabilidad, que representen en conjunto la segregación de todo el sistema. Para el caso continuo, y abusando de la notación, esto puede ser escrito por medio de una integral, como se muestra en (D.1). Para el caso, p representa percentiles, $\Lambda(p)$ el índice de segregación construido definiendo el ingreso de la observación en el percentil p como el límite de “vulnerabilidad”.

$$\Lambda^R = \sum_{j=1}^J \frac{t_j}{T} \left(\frac{v - v_j}{v} \right) = \int_0^1 w(p) \Lambda(p) dp \quad (\text{D.1})$$

El autor demuestra que, dependiendo de la definición de v y v_j , se pueden construir tres diferentes índices:

1. *Rank-order information theory index:*

$$H^R = 2 \log(2) \int_0^1 -(p \log_2(p) + (1 - p) \log_2(1 - p)) H(p) dp$$

2. *Rank-order variation ratio index:*

$$R^R = \frac{3}{2} \int_0^1 4p(1 - p) R(p) dp$$

3. *Rank-order square root index:*

$$S^R = \frac{4}{\pi} \int_0^1 2\sqrt{p(1 - p)} S(p) dp$$

$H(p)$, $R(p)$ y $S(p)$ son los índices de segregación teoría de la información, tasa de varianza y raíz cuadrada, construidos considerando el grupo con ingreso menor o igual al de la observación en el percentil p como “vulnerable”.

Como estas últimas funciones no se conocen, propone estimarlas a partir de los cálculos con ciertos percentiles. El autor recomienda usar valores de p que sean representativos, idealmente equidistantes. Para este trabajo se utilizan los cuantiles 0,1 a 0,9 con intervalos de 0,1. Se ajusta un polinomio de orden cinco¹ para obtener una aproximación

¹Mediante mínimos cuadrados ordinarios ponderados, para privilegiar la ajuste en las zonas medias. En base a los datos analizados en Reardon (2011), recomiendan utilizar polinomios de cuarto grado o mayores.

de las funciones. Por último, el autor propone una alternativa directa para el cálculo: si $\hat{B} = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$ es un vector de $(1 + k) * 1$ de los parámetros estimados (polinomio de grado k más la constante), entonces el índice resulta ser:

$$\Lambda^R = \Delta' \hat{B}$$

Donde $\Delta = (\delta_0, \delta_1, \delta_2, \dots, \delta_k)$ es un vector del mismo tamaño de $\hat{B}$, en que δ_m :

$$\delta_m = \begin{cases} \frac{2}{(m+2)^2} + 2 \sum_{n=0}^m \frac{(-1)^{m-n} \binom{m}{n}}{(m-n+2)^2} & \text{para } \Lambda^R = H^R \\ \frac{6}{(m+2)(m+3)} & \text{para } \Lambda^R = R^R \\ 4 \sum_{n=0}^m \frac{2n+1}{2n+4} & \text{para } \Lambda^R = S^R \end{cases} \quad (\text{D.2})$$

D.3 Descomposición

El índice propuesto admite descomposición, ya que su construcción es a partir de índices que lo permiten (Reardon, 2011). Si bien el autor no detalla procedimiento, siguiendo el ejemplo de Reardon (2009) para el caso del *ordinal segregation index*, se puede suponer el desarrollo. Considerando G grupos, y definiendo d_g y d como el nivel de diversidad para el grupo g y el sistema, respectivamente; se tiene que:

$$\begin{aligned} \Lambda^R &= \int_0^1 w(p) \Lambda(p) dp \\ &= \int_0^1 w(p) (\Lambda(p)_{\text{between}} + \Lambda(p)_{\text{within}}) dp \\ &= \int_0^1 w(p) \left(\Lambda(p)_{\text{between}} + \sum_{g=1}^G \frac{t_g d_g}{Td} \Lambda(p)_g \right) dp \\ &= \int_0^1 w(p) \Lambda(p)_{\text{between}} dp + \int_0^1 w(p) \sum_{g=1}^G \frac{t_g d_g}{Td} \Lambda(p)_g dp \end{aligned}$$

Estas integrales se resuelven de la misma manera propuesta en la sección anterior (y demostrada en Reardon, 2011):

$$\Lambda^R = \Delta' \hat{B}_{\text{between}} + \Delta' \hat{B}_{\text{within}}$$

Donde $\hat{B}_{\text{between}}$ es un vector de $(1 + k) * 1$ de los parámetros estimados del ajuste de un polinomio de grado k , cuya variable a ajustar es $\Lambda(p)_{\text{between}}$, para distintos valores de p (la segregación *between* para límites de vulnerabilidad fijados en el ingreso de la observación en el cuantil p). Y $\hat{B}_{\text{within}}$ es el mismo vector de parámetros ajustados, pero considerando $\sum_{g=1}^G \frac{t_g d_g}{T d} \Lambda(p)_g$ como la variable a ajustar (la segregación *within* para límites de vulnerabilidad fijados en el ingreso de la observación en el cuantil p). Δ representa lo mismo que en la sección anterior (ver ecuación D.2), y su construcción depende, al igual que $\Lambda(p)$, del índice que se quiera descomponer.

E. DEFINICIONES DE DISPONIBILIDAD

E.1 Criterios para Reducir Alternativas

Para operacionalizar los criterios de disponibilidad se define la variable $disp_{ij}$, que toma el valor 1 si el colegio j está disponible para el individuo i , 0 en caso contrario. Esta variable puede tomar valores intermedios, interpretados como una “probabilidad de disponibilidad”. Para estos casos, se define la disponibilidad de un establecimiento generando números aleatorios desde una distribución uniforme ($U(0, 1)$). Si $disp_{ij}$ es mayor a u , donde u es un número generado desde la distribución $U(0, 1)$, entonces j está disponible para i ¹.

E.1.1 Distancia

Un colegio j se considera disponible para un individuo i si las personas que más viajan para asistir al colegio j tienen que recorrer más de lo que tendría que trasladarse i , si el colegio al que asiste i ($j^*(i)$) está tan lejos como j , o si j está a menos de 2km. Es decir:

$$disp_{ij} = \begin{cases} 1 & \text{si } q(0,95, \{dist_{i'j} | i' \in A_j\}) \geq dist_{ij} \vee dist_{ij} \leq 1,05 \times dist_{ij^*(i)} \vee dist_{ij} \leq 2km \\ 0 & \text{e.o.c} \end{cases}$$

Donde $q(p, X)$ es el cuantil p del conjunto X , A_j es el conjunto de individuos que elige realmente la alternativa j , y $j^*(i)$ es la alternativa elegida por i .

E.1.2 Ingreso

La disponibilidad según ingreso está condicionada a percibir un ingreso al menos como el ingreso per cápita del percentil 5 de un conjunto de familias que asisten al colegio,

¹Esto implica que el conjunto de alternativas final varía según la semilla que se elija para generar los números aleatorios. Se estimó con distintas para evaluar impacto en parámetros, obteniendo cualitativamente los mismos resultados. En las simulaciones se aprovecha la “disponibilidad probabilística”, cambiando la semilla para cada repetición.

o a que el cobro de mensualidad sea menor o igual a 1,05 veces el pago declarado en el Cuestionario de Padres SIMCE 2006 (corregido para el caso de los particulares).

Para construir las distribuciones de ingreso por colegio se utilizaron las bases SIMCE 2003, 2006 y 2008; y Casen 2006 y 2009; ajustados por el índice de Ingresos mensual promedio per cápita según Encuesta Suplementaria de Ingresos, INE.

Se justificó dejar el percentil en vez del mínimo por la posibilidad de que el segundo reflejara un dato mal digitado (o mal declarado), al ver que numerosos colegios de mensualidad alta tenían familias con ingreso cercano a cero. Además, se podría suponer que parte de las observaciones con menores ingresos son estudiantes becados (ya sea por políticas del colegio, la ley del 15 % u otras medidas).

Solo se consideraron colegios para los que se tenía al menos 30 observaciones de ingreso (774 de 798). En la Figura E.1 se observa la diferencia entre la elección de diferentes percentiles como ingreso mínimo. Para evitar sesgos al truncar la distribución en el 5 %, se le asignó una probabilidad de disponibilidad a observaciones que no cumplían con el límite. Esta probabilidad se definió como la distribución acumulada empírica² evaluada en el ingreso de i ($\hat{F}_j(\text{ingreso}_i)$), más una función monótona estrictamente creciente que asegura continuidad ($f_j(\text{ingreso}_i)$)³.

$$\text{disp}_{ij} = \begin{cases} 1 & \text{si } \text{ingreso}_i \geq \text{ingresoPerc5}_j \vee \text{copago}_j \leq 1,05 \times \text{copago}_j \\ \hat{F}_j(\text{ingreso}_i) + f_j(\text{ingreso}_i) & \text{si } \text{ingreso}_i \geq \text{ingresoMin}_j \\ 0 & \text{si } \text{ingreso}_i < \text{ingresoMin}_j \end{cases}$$

A esto se agrega una restricción de ingreso máximo, con diseño similar a la de ingreso mínimo, solo que utilizando el máximo en vez de un percentil. Esta restricción se pudo

²Construida en base a las observaciones de ingreso del colegio j .

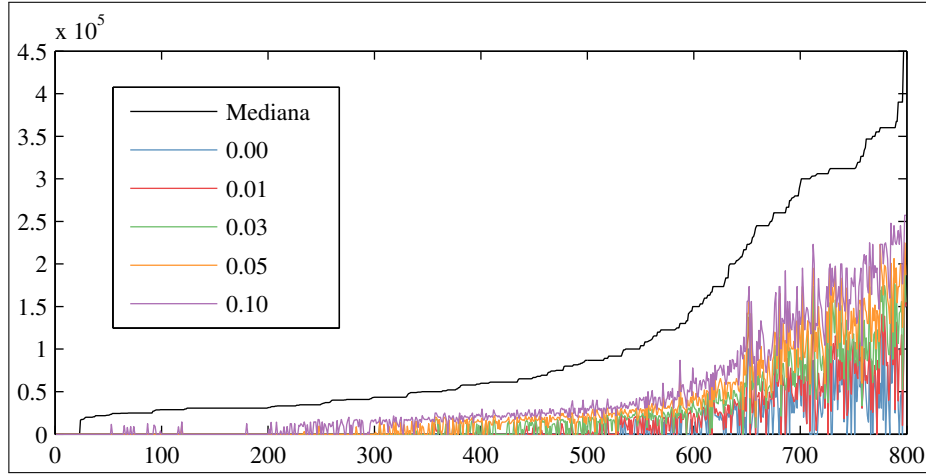
³La función permite que $\hat{F}_j(\text{ingresoPerc2}, 5_j) + f_j(\text{ingresoPerc2}, 5_j) = 1$ y cumple que $f_j(\text{ingresoMin}_j) = 0$. Se eligió una función basada en un comportamiento cuadrático.

construir solo para un 40 % de los colegios, pues el resto presentaba alguna observación de ingreso truncado⁴, criterio suficiente para considerar información incompleta.

$$disp_{ij} = \begin{cases} 1 & \text{si } ingreso_i \leq ingresoMax_j \\ 0 & \text{si } ingreso_i > ingresoMax_j \end{cases}$$

Estas restricciones empíricas se basan en el supuesto que la composición socioeconómica de los colegios no varía de gran manera, y que los padres consideran los colegios en que, al menos en la zonas extremas de la distribución, atienden familias del mismo poder adquisitivo. Se agrega la condición de pago de mensualidad para las excepciones en que las familias están dispuestas a gastar un mayor porcentaje de sus ingresos en educación, lo que puede no ser bien reflejado en las distribuciones de ingreso construidas para fijar el mínimo de ingreso por colegio, o estudiantes que acceden a becas.

FIGURA E.1. Ingresos per cápita de familias por colegio (ordenados por mediana)



E.1.3 Timing de elección

Definiendo $EmpiezaEn7mo_j = 1$ si el colegio j tiene cursos solo a partir de séptimo básico o un nivel superior; $SoloTP_j = 1$ si el único tipo de enseñanza que ofrece en

⁴Esto se debe a que las declaraciones de ingreso en cuestionarios SIMCE son categóricas.

media es técnico profesional; $NoSeCambia_i = 1$ si el individuo i no se cambia de colegio después de sexto básico; y $private_{j^*(i)} = 1$ si el colegio al que asiste i es privado, entonces la disponibilidad condicionada al *timing* de elección es:

$$disp_{ij} = \begin{cases} 0 & \text{si } (EmpiezaEn7mo_j = 1 \vee SoloTP_j = 1) \wedge (NoSeCambia_i = 1 \wedge private_{j^*(i)} = 1) \\ 1 & \text{e.o.c.} \end{cases}$$

Esta restricción se basa en que familias que asisten a colegios privados tienen una tendencia mucho menor al cambio que en otras dependencia, lo que podría significar que para muchos casos no evalúen colegios que empiecen en educación media, o un poco antes, en la decisión de elección de colegio. Para todos aquellos que sí realizan un cambio, independiente si se mantienen en el sector privado, sí se les deja disponibles colegios de solo media.

E.1.4 Notas

La construcción del criterio de selección académica se basa en distribuciones empíricas de las “notas de presentación” de las observaciones que se cambian de colegio. Estas últimas se definen como el promedio general del año anterior al cambio de colegio⁵. En el anexo F se realiza un pequeño análisis, y se explica qué colegios clasifican como los que seleccionan por nota.

En vista que esta selección no necesariamente se traduce en un promedio general mínimo de entrada⁶, sino que también puede ser un examen de admisión u notas específicas a ciertos años o materias, se procedió de manera similar a la restricción por ingreso. Para los colegios que se considera que seleccionan, se fijó el valor de la moda de las notas de presentación de todas las observaciones que se cambiaron al colegio j como la cota

⁵Para el caso de que no se tenga el año inmediatamente anterior se buscó uno antecesor. Si no existe, entonces se imputó como el máximo (conservador) de las notas de presentación de sus compañeros del colegio nuevo con promedio similar.

⁶Inspección visual de las distribuciones de notas de los matriculados en los colegios que supuestamente seleccionan, no muestran ninguna evidencia de un corte importante a la izquierda, ni siquiera en los establecimientos más exigentes.

inferior que asegura disponibilidad. Los individuos con nota inferior a ésta, y superior a la nota mínima observada, tienen una probabilidad de disponibilidad asociada a la distribución acumulada empírica ($\hat{F}_j(nota_i)$)⁷, y una función monótona estrictamente creciente ($f_j(nota_i)$). El criterio de disponibilidad se resume en lo siguiente:

$$disp_{ij} = \begin{cases} 1 & si \quad nota_i \geq notaModa_j \vee (NoSeCambia = 1 \wedge private_{j^*(i)} = 1) \\ \hat{F}_j(nota_i) + f_j(nota_i) & si \quad nota_i \geq notaMin_j \\ 0 & si \quad nota_i < notaMin_j \end{cases}$$

En la Figura E.2 se observan tres ejemplos, cada columna corresponde a un colegio. Los gráficos de la primera fila corresponden a los histogramas de las notas de todos los individuos que se cambian en la realidad. La segunda fila corresponde a la distribución de las notas de los individuos que tienen disponible el colegio, según la metodología recién propuesta. La tercera fila es la misma distribución anterior, si se estableciera disponible el colegio a todas las observaciones mayores a la nota mínima de la distribución real (fila 1). Por último, la cuarta columna es similar a la tercera, pero fijando el límite de disponibilidad en la nota del percentil cinco en vez de la mínima.

Se aprecia que el universo de individuos con colegios disponible según el método propuesto imita mucho mejor la distribución real, lo que hace más probable que la distribución de notas de los alumnos que según el modelo eligen el colegio también lo haga. Por consiguiente, se acerca más a la selección por nota que realizan los establecimientos, a si se utilizara otro criterio.

E.1.5 Co-educacional

Criterio bastante simple, un colegio j no está disponible si solo recibe a un género ($SoloMujeres_j = 1$ o $SoloHombres_j = 1$) si i es del sexo opuesto ($DFemale_i = 0$ o

⁷Construida en base a las notas de presentación de todas las observaciones que se cambian al colegio j .

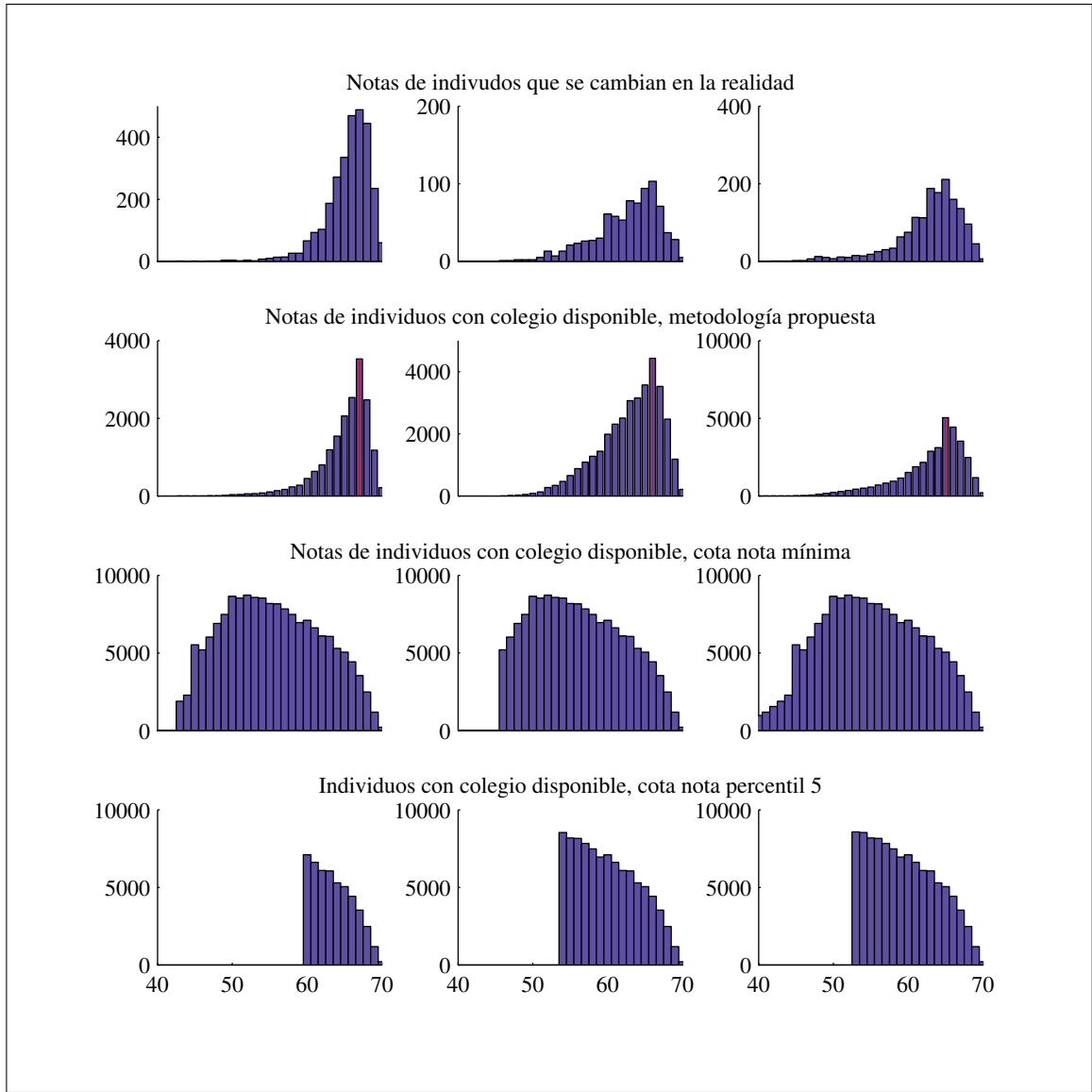


FIGURA E.2. Ejemplo de restricción de disponibilidad por nota

$DFemale_i = 1$, respectivamente).

$$disp_{ij} = \begin{cases} 0 & \text{si } (DFemale_i = 0 \wedge SoloMujeres_j = 1) \vee (DFemale_i = 1 \wedge SoloHombres_j = 1) \\ 1 & \text{e.o.c.} \end{cases}$$

E.2 Disponibilidad Geográfica

Se procedió a calcular la cantidad de colegios disponibles según dependencia en un radio delimitado (definido en 2km) junto con las características promedio de estos colegios. La motivación detrás de este análisis es la gran varianza en atributos que existe entre los colegios, incluso dentro de colegios de la misma dependencia. Se vuelve interesante comparar los atributos de la oferta (definida solamente en términos de cercanía) según nivel socioeconómico. La Tabla E.1 detalla los resultados de los cálculos sin considerar las restricciones de disponibilidad presentados anteriormente. Si se considera las dependencias municipales y subvencionados, el SIMCE de los colegios que rodean a los más “vulnerables” es menor y crece montónicamente con el ingreso (al igual que el el copago). El numero de colegios municipales promedio disponibles para los ocho primeros deciles es bastante parejo, variando de 2,7 a 3,2. Este numero se eleva para el noveno decil, con 3,4 y cae para el décimo, con 2 colegios en el radio de 2km. Todos enfrentan colegios municipales tamaño similar, salvo el decil más afluente que decae cerca de un 20 %. La cantidad de colegios subvencionados no varía en gran manera, enfrentando casi todos cerca de 10 colegios, con excepción del grupo con más recursos que el número baja a 3,7. Por último, la cantidad de colegios privados promedio en el entorno de familias de los siete primeros deciles es muy baja, variando entre 1,0 y 2,3, mientras que para los tres grupos de mayor ingreso sube a 3,5; 7,2 y 9,6.

Cuando se consideran las restricciones de disponibilidad impuestas en este trabajo (Tabla E.2), se aprecia que baja de manera importante la disponibilidad de colegios subvencionados para el decil más pobre (de 9,6 a 7,2) y con esto el SIMCE promedio. El gran impacto, por cierto esperable, es en la cantidad de colegios privados en el entorno, que pasa de 1 a 0,3 para el 30 % más “vulnerable”, con reducciones considerables para el 30 % siguiente también. Al imponer las restricciones aumenta la diferencia en rendimiento entre los colegios en las cercanías de los más pobres y de los más ricos, independiente de la dependencia que se analice.

TABLA E.1. Colegios disponibles en radio según decil, sin disponibilidad restringida

Decil	Municipal					Subvencionado					Particular				
	Cantidad	Tamaño	SIMCE	Copago	Católico	Cantidad	Tamaño	SIMCE	Copago	Católico	Cantidad	Tamaño	SIMCE	Copago	Católico
(1)	2.9	16.0	225.4	0.6	0.0	9.6	10.2	248.1	18.2	0.3	1.0	4.7	304.1	129.4	0.2
(2)	2.8	16.2	226.2	0.6	0.0	10.1	10.2	248.9	18.6	0.3	1.0	4.6	304.7	128.5	0.2
(3)	2.7	16.1	227.6	0.6	0.0	10.2	10.2	249.2	19.2	0.3	1.1	4.5	304.8	128.6	0.2
(4)	2.9	16.1	229.2	0.6	0.0	10.6	10.3	249.8	19.9	0.3	1.3	4.6	304.9	128.7	0.3
(5)	2.8	15.9	230.9	0.7	0.0	11.0	10.3	251.4	21.0	0.3	1.4	4.4	304.5	130.1	0.3
(6)	3.0	16.5	231.6	0.7	0.0	11.1	10.5	251.6	21.6	0.3	1.7	4.5	303.9	131.5	0.3
(7)	3.1	16.2	233.0	1.0	0.0	11.5	10.3	254.0	23.0	0.2	2.3	4.7	302.8	140.9	0.3
(8)	3.2	16.2	235.3	1.4	0.0	11.3	10.3	256.4	25.1	0.2	3.5	5.0	302.2	155.9	0.3
(9)	3.4	15.7	242.3	2.7	0.0	9.2	9.8	261.6	28.1	0.3	7.2	5.2	302.9	172.0	0.3
(10)	2.0	13.3	251.2	5.5	0.0	3.7	8.4	271.0	30.2	0.4	9.6	6.0	310.0	212.3	0.4

Copago: mensualidad expresada en miles de pesos.

TABLA E.2. Colegios disponibles en radio según decil, con disponibilidad restringida

Decil	Municipal					Subvencionado					Particular				
	Cantidad	Tamaño	SIMCE	Copago	Católico	Cantidad	Tamaño	SIMCE	Copago	Católico	Cantidad	Tamaño	SIMCE	Copago	Católico
(1)	2.5	15.6	221.1	0.6	0.0	7.2	10.5	240.2	14.9	0.2	0.2	2.8	296.0	124.3	0.3
(2)	2.4	15.6	221.6	0.5	0.0	8.4	10.1	243.8	17.2	0.2	0.3	3.6	295.4	121.2	0.3
(3)	2.4	15.8	223.0	0.5	0.0	8.7	10.2	244.8	17.9	0.2	0.3	4.6	297.7	124.8	0.3
(4)	2.5	15.8	223.7	0.6	0.0	9.0	10.2	245.7	19.2	0.2	0.5	4.3	299.0	125.7	0.3
(5)	2.4	15.4	224.7	0.6	0.0	9.4	10.2	247.2	20.4	0.2	0.8	4.4	298.8	129.8	0.3
(6)	2.5	15.8	225.4	0.7	0.0	9.5	10.3	247.6	20.9	0.2	1.1	4.6	299.0	137.5	0.3
(7)	2.5	15.8	227.4	0.8	0.0	9.8	10.3	250.2	22.6	0.2	1.7	4.6	299.6	148.5	0.3
(8)	2.4	15.9	228.9	1.0	0.0	9.4	10.4	253.5	25.0	0.2	2.8	4.8	300.4	162.5	0.3
(9)	1.8	15.9	235.1	1.9	0.0	6.9	10.0	260.1	28.9	0.2	6.0	5.3	302.6	183.3	0.4
(10)	0.8	13.3	254.2	6.0	0.0	2.6	8.1	273.9	32.9	0.3	8.1	6.0	309.7	213.6	0.4

Copago: mensualidad expresada en miles de pesos.

E.3 Conjunto de Alternativas Aleatorio

En modelación de procesos de elección discreta en que el número de alternativas es muy grande, tradicionalmente se podía optar por acotar las alternativas disponibles de manera aleatoria; sin introducir sesgo en los parámetros Ortúzar y Willumsen (2011). Si

bien esto hace sentido pensando en lo poco realista que los padres consideren realmente más de 100 alternativas (o más de 5!), optar por disponibilidad aleatoria no aporta en estimar parámetros menos sesgados, pues la eliminación de alternativas no está fundada en reales consideraciones de los electores. El gran beneficio es bajar la carga computacional asociada a la consideración de todas las alternativas; pero Nerella y Bhat (2004) demuestra que en modelos más complejos que el MNL, como lo son *mixed logit* o *latent class logit*, esta estrategia es potencial fuente de sesgo. Por las razones anteriores no se considera la reducción de alternativas con criterio aleatorio; además que la precisión de los software actuales, y la capacidad de cálculo, permiten que no sea restrictiva la estimación con todas las alternativas⁸.

⁸Se utilizaron dos computadores personales para la estimación, con procesadores doble núcleo Intel Core i7 de 2,7 y 2,8 GHz respectivamente, 8GB de memoria RAM, y el programa Matlab 7 64-bit; uno de ellos fue facilitado por Giorgio Jackson, gestión de Camila Ponce. Matriz de covarianzas se calculó en un servidor del Departamento de Ingeniería en Transporte UC (Intel Xeon 3,4 GHz quad core, 16GB de memoria RAM); agradecimientos al profesor Juan de Dios Ortúzar, Ignacia Torres y Edgar Tembladera por sus gestiones.

F. SELECCIÓN ACADÉMICA

No se encontraron estudios con evidencia empírica convincente relacionados con los proceso de selección académica, en colegios enfocados en la educación media. Carrasco et al. (2014) se aproxima a la selección a través de entrevistas a directores, concentrado principalmente en educación básica. Un resultado interesante que llegan los autores es que la selectividad no es exclusiva de los colegios con alta demanda (más postulantes que cupos), en especial en colegios particulares subvencionados, más bien es una práctica común para poder competir. Valenzuela y Allende (2012), por otra lado, buscan encontrar las diferencias en valor agregado de Liceos Públicos de Excelencia, realizando *matching* en base a *propensity score*. Los autores modelan la selección académica de forma indirecta, utilizando un *proxy* de habilidad a partir el residuo individual de una función de producción de SIMCE de 4to básico. Concluyen que la gran diferencia en resultados se reduce considerablemente cuando se controla por habilidad anterior y características socioeconómicas.

En vista de la escasez, y para sustentar las restricciones de selección académica, se realizó un análisis del cambio de colegio¹ y nivel académico de los postulantes a diferentes colegios.

F.1 Movilidad

De las 37.859 observaciones utilizadas en la estimación, un 67 % se cambia de colegio entre sexto básico y segundo medio. El 56 % de estos cambios son a un colegios de solo media (o de séptimo en adelante). En la Tabla F.1 se reporta el flujo de alumnos entre dependencias. Se aprecia que los cambios hacia colegios municipales están bastante bien repartidos entre alumnos de colegios subvencionados como públicos, mientras que el flujo hacia los subvencionados es dominado por alumnos que provienen de establecimientos de la misma dependencia. Tres cuartos de la observaciones que se cambian a un colegio privado originalmente venían de otro del mismo tipo, mientras que el 20 % estaba en un

¹Un estudio detallado de la movilidad escolar se puede ver en Larroulet (2011).

subvencionado y el 6 % en un municipal. Es interesante que la proporción de observaciones que se cambian de privado a privado es solo levemente superior a los que lo hacen de privado a subvencionado: esto impacta en que el neto para el sistema privado sea negativo.

En la última fila se describe el porcentaje de la matrícula de segundos medios 2006 que se explica por cambios de colegios. Gran parte de los individuos de los colegios municipales son “nuevos”, en el sentido que no estaban en el mismo colegio antes de séptimo. Sigue siendo considerablemente alta esta proporción para el sistema subvencionado, con sobre dos tercios; bajando bruscamente a menos de un 30 % para el caso de los privados.

TABLA F.1. Flujo de cambios de colegio entre dependencias

Desde	Hacia			
	Municipal	Subvencionado	Privado	Todos
Municipal	4479 (47 %)	4488 (32 %)	96 (6 %)	9063 (36 %)
Subvencionado	4648 (49 %)	8590 (60 %)	357 (21 %)	13595 (53 %)
Privado	398 (4 %)	1138 (8 %)	1284 (74 %)	2820 (11 %)
Todos	9525 (100 %)	14216 (100 %)	1737 (100 %)	25478 (100 %)
% Matrícula	90 %	67 %	28 %	67 %

La Tabla F.2 muestra algunos estadísticos referidos a los cambios de colegio desagregados por decil de ingreso. Se observa que las notas de presentación² promedio presentan una suave tendencia a crecer a medida que aumenta el ingreso. Un dato que tiene gran varianza es la cantidad de individuos que se cambian de colegio según nivel de ingreso, disminuyendo casi linealmente entre el primer y octavo decil (entre un 82 % y un 65 %), y fuertemente para los dos más afluentes; donde se cambian de colegio un 51 % y un 26 % de las observaciones.

²Nota de presentación se definió como el promedio general del año anterior al cambio de colegio. Para el caso de que no se tenga el año inmediatamente anterior se buscó uno antecesor. Si no existe, entonces se imputó como el máximo (conservador) de las notas de presentación de sus compañeros del colegio nuevo con promedio similar. A las observaciones que no realizan cambio de colegio se les definió como la nota del año más antiguo presente en la base (para la mayoría sexto básico).

Las medias de las notas de los que se cambian son iguales a las del grupos, menos considerando los dos más afluentes, lo que puede está relacionado a los diferentes motivos de cambio de colegio predominantes según ingreso (no desarrollado en este trabajo). El porcentaje de estudiantes que se cambian a un colegio municipal disminuye con el ingreso, mientras que aumenta para el caso de los subvencionados. Prácticamente nadie de los seis primeros deciles se cambia a un colegio privado, en contraste con el 81 % del decil con más recursos.

TABLA F.2. Cambios de colegio según decil de ingreso per cápita

	Decil									
	1	2	3	4	5	6	7	8	9	10
Nota promedio	5.6	5.7	5.8	5.8	5.8	5.8	5.9	5.9	5.9	6.0
% se cambia	82	80	79	77	75	72	68	65	51	26
Nota promedio se cambia	5.6	5.7	5.8	5.8	5.8	5.9	5.9	5.9	5.8	5.8
% se cambia a municipal	48	44	42	42	41	39	37	31	18	4
% se cambia a subvencionado	52	56	58	57	60	60	62	62	49	15
% se cambia a privado	0	0	0	0	0	1	3	8	33	81

F.2 Perfil Académico de Nuevos Estudiantes

El siguiente ejercicio consiste en agrupar las notas de presentación de las observaciones que se cambian de colegio³ para tener una noción de qué tipo de estudiante recibe cada establecimiento. Para todos las Figuras expuestas a continuación, los primeros *box-plots*⁴ corresponden a las distribuciones de notas de todas las observaciones que se cambian de colegio separados por la dependencia a la que se cambian; donde amarillo representa a todos, azul municipales, rojo subvencionados y verde privados.

³Todas las observaciones que se cambian entre los años 2002 y 2006, no solo los de la base de estimación. Solo se consideran establecimientos para los que se cambien al menos 40 estudiantes.

⁴La caja corresponde al rango intercuartil, las líneas que se extienden fuera de las cajas representan el primer y último decil. El asterisco es la media y la línea dentro de la caja la mediana.

Los otros *box-plots*, a la derecha de la línea punteada gris, corresponden a las distribuciones de los 15 mejores (o peores) colegios según SIMCE⁵. Para evitar potencial confusión con selección socioeconómica, presumiblemente correlacionada con copago; se analizó separadamente según cobro⁶. Se definieron cuatro tramos de cobro: menor igual a \$7.500, a \$17.500, a \$60.000 y superior a \$60.000. Los resultados se muestra en las Figuras de la F.1 a la F.8.

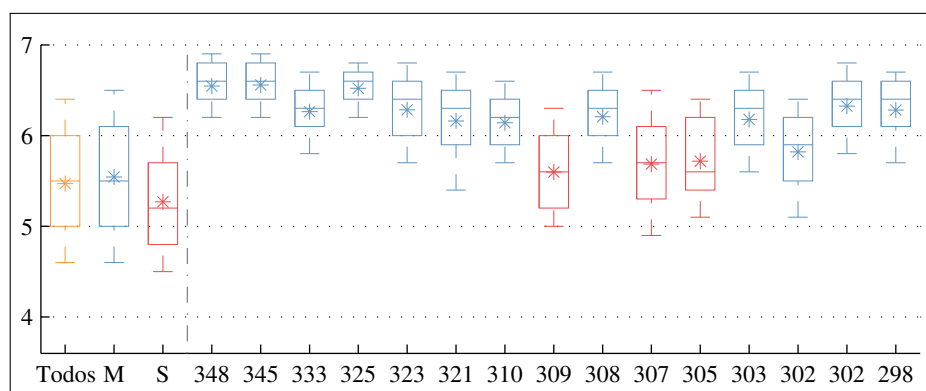


FIGURA F.1. Distribución de notas (eje Y) en colegios con mejor SIMCE (eje X) con cobro $\leq$ \$7.500

Se observa que, independiente de la dependencia, los colegios con mejor rendimiento reciben alumnos con mejor nota de presentación. Este fenómeno es más claro en la Figura F.1. Los colegios “buenos y baratos” son aparentemente muy selectivos, efecto más evidente en los municipales que en subvencionados⁷. En los dos establecimientos con mejores resultados, casi tres cuartos de sus alumnos matriculados son del 10 % de mejor nota de los aspirantes a colegios de esta categoría de cobro.

⁵Se consideraron únicamente los colegios a los que llegan más de 40 estudiantes nuevos en el período 2002-2006

⁶Esto implica que los *box-plots* a la izquierda de la línea punteada gris de las Figuras son la distribución de notas de todos los que se cambian a los colegios dentro del rango de cobro analizado.

⁷Cabe destacar las tres excepciones de la Figura F.1, todos colegios subvencionados con rendimientos cercanos establecimientos municipales, que reciben alumnos de “peor” historial académico.

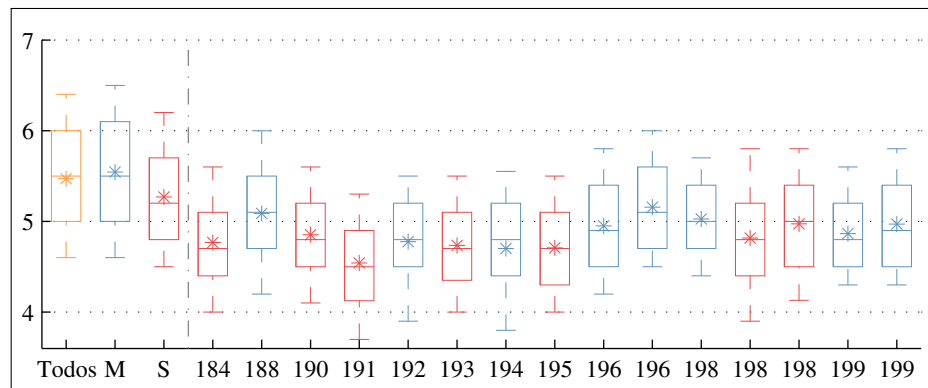


FIGURA F.2. Distribución de notas (eje Y) en colegios con peor SIMCE (eje X) con cobro $\leq$ \$7.500

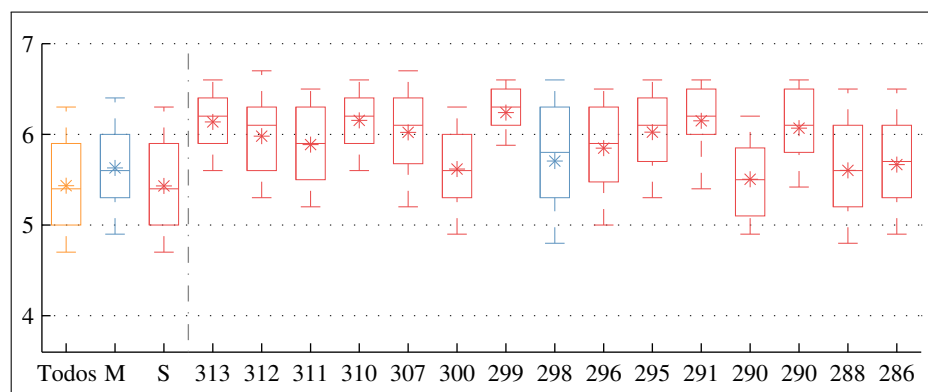


FIGURA F.3. Distribución de notas (eje Y) en colegios con mejor SIMCE (eje X) con cobro $\leq$ \$17.500

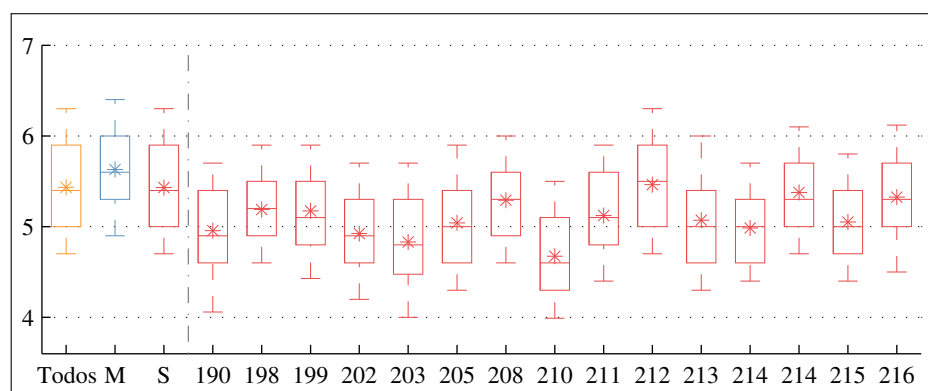


FIGURA F.4. Distribución de notas (eje Y) en colegios con peor SIMCE (eje X) con cobro $\leq$ \$17.500

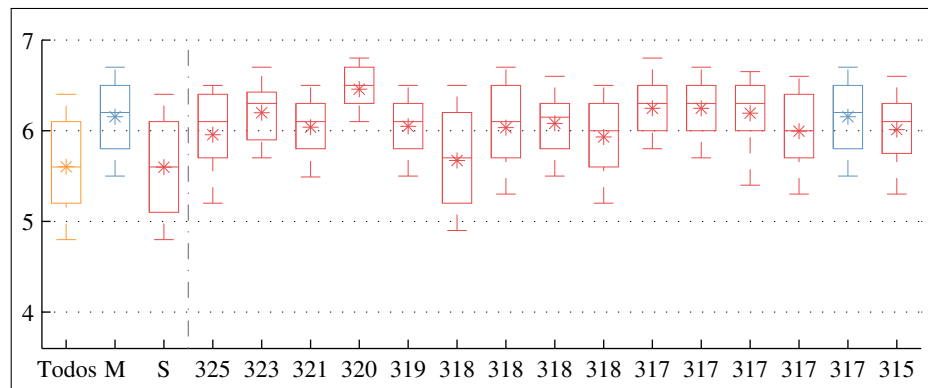


FIGURA F.5. Distribución de notas (eje Y) en colegios con mejor SIMCE (eje X) con cobro $\leq \$60,000$

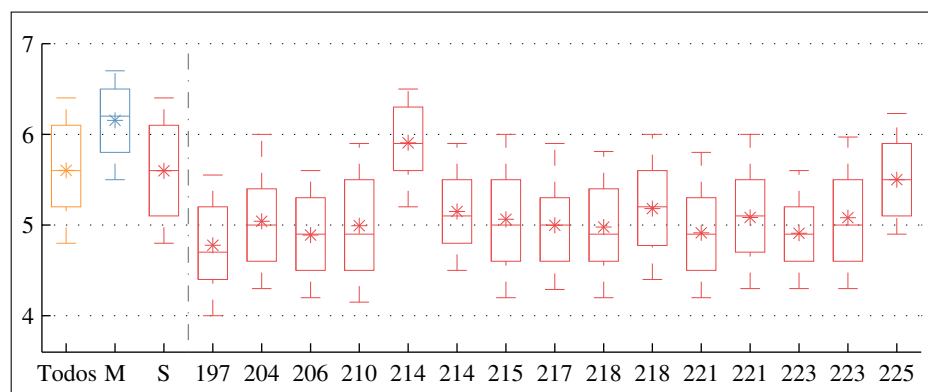


FIGURA F.6. Distribución de notas (eje Y) en colegios con peor SIMCE (eje X) con cobro $\leq \$60,000$

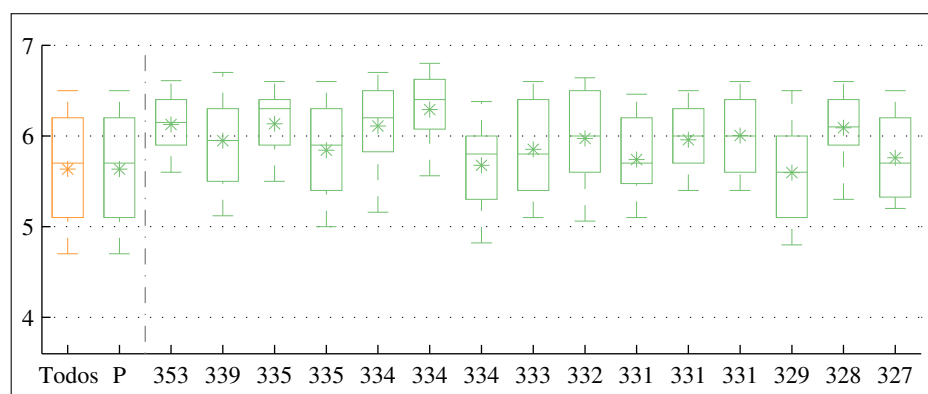


FIGURA F.7. Distribución de notas (eje Y) en colegios con mejor SIMCE (eje X) con cobro $> \$60,000$

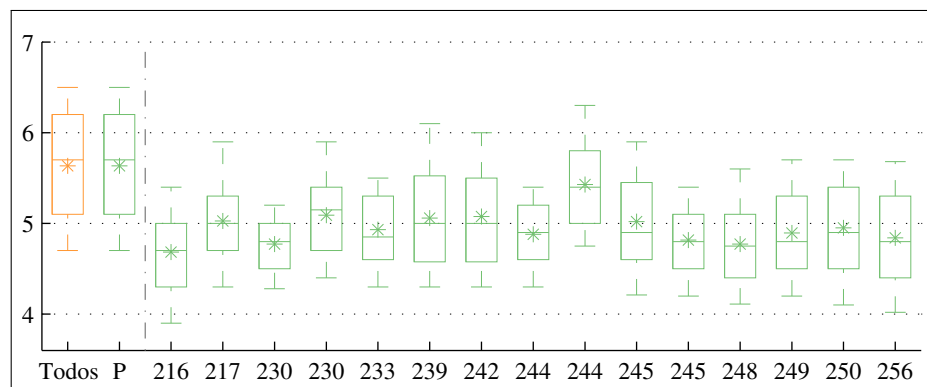


FIGURA F.8. Distribución de notas (eje Y) en colegios con peor SIMCE (eje X) con cobro > \$60.000

El fenómeno opuesto se observa en los colegios con rendimiento más bajo (Figura F.2 y otras); en general, reciben alumnos con resultados académicos bastante menores al promedio, donde nuevamente los municipales lo hacen peor, recibiendo mejores alumnos tienen rendimientos similares a los subvencionados (solo por inspección visual). Al parecer, así como existen hipótesis se que hay “colegios para cada nivel socioeconómico”, también puede que existan “colegios para cada nivel de rendimiento”.

De los gráficos F.7 y F.8 se observa que este fenómeno no es exclusivo de los colegios con aporte estatal; en establecimientos privados existe también correlación entre SIMCE y nota de los alumnos que aceptan.

F.3 Colegios que Seleccionan

Se definió que un colegio puede potencialmente realizar selección académica si la media de notas del total de alumnos que se cambian a ese colegio es estadísticamente superior a la media de las notas todas las observaciones que se cambian de colegio⁸. Se siguió esta estrategia pues no se encontraron cortes bruscos en las distribuciones de notas

⁸Considerando todas las observaciones y considerando solo las que se cambian a colegios de nivel de precio similar, en definitiva es necesario pasar dos tests. Se tomaron en cuenta todos los cambios entre los años 2002 y 2006, independiente del curso al que se cambie, y un nivel de significancia del 1 % ($H_o : \mu = \mu_j$, $H_a : \mu < \mu_j$). Solo se evaluó la selección para colegios en que hay al menos 30 cambios.

de los estudiantes que se cambian a cada colegio, incluso en los establecimientos más exigentes.

Bajo este criterio, existen 255 establecimientos para los que la admisión es condicional a la nota que se obtuvo en el colegio anterior, lo que equivale a un 32 % del sistema. La Tabla F.3, muestra porcentaje de colegio que realizan selección académica⁹ según dependencia y cobro¹⁰ ; junto con los puntajes SIMCE promedio de los colegios con y sin selección. Se observa que colegios que seleccionan tienen puntaje más alto, diferencia que va entre 15 y 51 puntos¹¹.

TABLA F.3. Colegios que seleccionan por notas y relación con SIMCE

Cobro	% Selecciona			SIMCE Seleccionan			SIMCE No seleccionan		
	M	S	P	M	S	P	M	S	P
Bajo	31	22	17	273	270	314	222	229	282
Medio	75	33	37	277	276	321	278	246	296
Alto	-	49	32	-	298	329	-	259	314
Todos	33	33	29	274	283	322	222	241	296

⁹Esta selección es muy heterogénea. Los “notas mínimas” para los que se permite la entrada son variables según colegio. Esto es algo que se tiene muy considerado en la modelación.

¹⁰Cobro bajo para municipales y subvencionados es menor o igual a \$12.500, medio menor o igual a 27.500. Para privados, estas cifras son \$140.000 y \$187.000 respectivamente.

¹¹Sin contar los colegios municipales de copago medio, que siendo solo cuatro observaciones no difieren en rendimiento.

G. ESTIMACIÓN

G.1 Primera Etapa: *EM algorithm*

El *EM algorithm* fue por primera vez introducido en Dempster et al. (1977), y es una estrategia para resolver de manera iterativa modelos que se vuelven complejos con métodos de maximización tradicionales (Greene, 2001; Train, 2009). Originalmente fue utilizado para tratar valores omitidos en bases de datos, pero su uso se extendió a más áreas (McLachlan y Krishnan, 2007). Siguiendo la notación de Greene (2001), su aplicación en estimación de modelos de clase latente es la siguiente. Definiendo $u_{is} = 1$ si el individuo i pertenece a la clase s , $u_{is} = 0$ en caso contrario, la función de verosimilitud (G.1) se puede escribir como (G.2), o su versión en logaritmo (G.3).

$$\mathcal{L}(\theta, \lambda \mid Z, W, M) = \prod_{j=1}^J \prod_{i \in A_j} \sum_{s=1}^S P_{is} P_{ij|s} \quad (\text{G.1})$$

$$= \prod_{j=1}^J \prod_{i \in A_j} \prod_{s=1}^S (P_{ij|s})^{u_{is}} \quad (\text{G.2})$$

$$\ell(\theta, \lambda \mid Z, W, M) = \sum_{j=1}^J \sum_{i \in A_j} \sum_{s=1}^S u_{is} \log(P_{ij|s}) \quad (\text{G.3})$$

El paso de esperanza (*expectation step*) implica reemplazar los valores omitidos (en este caso la pertenencia a las clase: u_{is}) por su esperanza condicional en las características y elección de los individuos (notar que no es lo mismo que P_{is} , que sería la esperanza de la pertenencia a la clase s condicional únicamente en las características M_i de individuo, no en la elección). En el paso de maximización (*maximization step*) se busca máximo para (G.3) condicional a estas esperanzas de pertenencia (es decir asumiéndolas constante), lo que se traduce en estimar tantos logits (en su versión tradicional: logit condicional o MNL) como numero de clases, con las observaciones ponderadas por $E(u_{is})$. También se

actualiza la probabilidad de pertenencia a la clase condicional en las características de los individuos (P_{is}).

La esperanza condicional de u_{is} descrita en (G.4)¹ se obtiene a través del teorema de Bayes. Para el cálculo de $P_{ij|s}$ y P_{is} , se utilizan los parámetros estimados en la iteración anterior λ y θ ².

$$E(u_{is}) = \frac{P_{is}P_{ij^*|s}}{\sum_{s'=1}^S P_{is'}P_{ij^*|s'}} \quad (\text{G.4})$$

El proceso es iterativo en el sentido que utiliza los parámetros (λ^t y θ^t) estimados en una iteración para calcular los pesos (*expectation step*: $E(u_{is})^{t+1} = f(\lambda^t, \theta^t)$) y recalculan los parámetros en función de éstos (*maximization step*: $[\lambda^{t+1}, \theta^{t+1}] = g(E(u_{is})^{t+1})$). McLachlan y Krishnan (2007); Wu (1983) demuestran que el proceso es bastante estable y que luego de repetidas iteraciones converge al máximo de (G.1)³. Una de las críticas que recibe es su velocidad de convergencia (Bhat, 1997; Greene y Hensher, 2003), por lo que se sugiere utilizar estrategias mixtas, como la que se desarrolla en el presente trabajo.

G.2 Segunda Etapa: Optimización Directa

Utilizando el resultado de la primera etapa como valores iniciales, se optimizó la función objetivo mediante el algoritmo de maximización BHHH, implementado en Matlab⁴.

Primero, se separan las características de las alternativas (W) en B , como las percibidas por cada individuos de forma distinta; y X , como las percibidas por todos igual.

¹ j^* representa la alternativa elegida por i . Estrictamente debiera ser $j^*(i)$, pero se omite el paréntesis.

² λ es el resultado de la estimación de S modelos logit con observaciones ponderadas por $E(u_{is})$, mientras que θ es el resultado de la estimación de un modelo logit multinomial de pertenencia a clases según las características de segmentación (M), que como variable dependiente utiliza $E(u_{is})$.

³Los autores son enfáticos que esto sucede “bajo condiciones normales”.

⁴Basado en código para maximización de función de verosimilitud mediante BHHH (`maxlike.m`) de Thomas H. Jørgensen, estudiante de doctorado University of Copenhagen, (www.econ.ku.dk/phdstudent/jorgensen/code.htm).

Luego, y asumiendo heterogeneidad en preferencias según características de los individuos (Z); se puede reescribir la probabilidad de elección condicional a la pertenencia a una clase $P_{ij|s}$, y la probabilidad de pertenencia a la clase P_{is} , como se muestra en (G.5) y (G.6). La función a maximizar (G.7) es el logaritmo de la probabilidad conjunta de las elecciones reales (función de log-verosimilitud).

$$P_{ij|s} = \frac{\exp \left(\sum_{r=1}^R \sum_{h=1}^H \gamma_{rhs} B_{ijh} Z_{ir} + \sum_{r=1}^R \sum_{k=1}^K \beta_{rks} X_{jk} Z_{ir} \right)}{\sum_{\forall j' \in D_i} \exp \left(\sum_{r=1}^R \sum_{h=1}^H \gamma_{rhs} B_{ij'h} Z_{ir} + \sum_{r=1}^R \sum_{k=1}^K \beta_{rks} X_{j'k} Z_{ir} \right)} \quad (\text{G.5})$$

$$P_{is} = \frac{\exp \left(\sum_{t=1}^T M_{it} \theta_{ts} \right)}{\sum_{s'=1}^S \exp \left(\sum_{t=1}^T M_{it} \theta_{ts'} \right)} \quad (\text{G.6})$$

$$\begin{aligned} \begin{bmatrix} \hat{\gamma} \\ \hat{\beta} \\ \hat{\theta} \end{bmatrix} &= \arg \max_{\gamma \beta \theta} \sum_{j=1}^J \sum_{i \in A_j} \log \left(\sum_{s=1}^S P_{is} P_{ij|s} \right) \\ &= \arg \max_{\gamma \beta \theta} \ell \end{aligned} \quad (\text{G.7})$$

G.2.1 Condiciones de primer orden

Las condiciones de primer orden son las primeras derivadas de la función con respecto a los parámetros a estimar (jacobiano de la función de log-verosimilitud):

$$\begin{aligned} \frac{\partial \ell}{\partial \gamma_{rhs}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{P_{is} P_{ij|s}}{\sum_{s'=1}^S P_{is'} P_{ij|s'}} \left(B_{ijh} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} B_{ij'h} Z_{ir} \right) \\ \frac{\partial \ell}{\partial \beta_{rks}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{P_{is} P_{ij|s}}{\sum_{s'=1}^S P_{is'} P_{ij|s'}} \left(X_{jk} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} X_{j'k} Z_{ir} \right) \\ \frac{\partial \ell}{\partial \theta_{ts}} &= \sum_{j=1}^J \sum_{i \in A_j} M_{ts} \left(\frac{P_{is} P_{ij|s}}{\sum_{s'=1}^S P_{is'} P_{ij|s'}} - P_{is} \right) \end{aligned}$$

G.2.2 Condiciones de segundo orden

Las condiciones de segundo orden corresponden a las segundas derivadas de la función con respecto a los parámetros a estimar. Por temas de espacio, se utilizará la notación de la esperanza de la pertenencia a clases condicional a características del individuo y de la elección definida en (G.4):

$$\begin{aligned}
\frac{\partial^2 \ell^2}{\partial \gamma_{rhs} \partial \gamma_{\check{r}\check{h}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \gamma_{\check{r}\check{h}\check{s}}} \left(B_{ijh} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} B_{ij'h} Z_{ir} \right) - E(u_{is}) \left(\sum_{\forall j' \in D_i} \frac{\partial P_{ij'|s}}{\partial \gamma_{\check{r}\check{h}\check{s}}} B_{ij'h} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \gamma_{rhs} \partial \beta_{\check{r}\check{k}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \beta_{\check{r}\check{k}\check{s}}} \left(B_{ijh} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} B_{ij'h} Z_{ir} \right) - E(u_{is}) \left(\sum_{\forall j' \in D_i} \frac{\partial P_{ij'|s}}{\partial \beta_{\check{r}\check{k}\check{s}}} B_{ij'h} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \gamma_{rhs} \partial \theta_{\check{t}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \theta_{\check{t}\check{s}}} \left(B_{ijh} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} B_{ij'h} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \beta_{rhs} \partial \gamma_{\check{r}\check{h}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \gamma_{\check{r}\check{h}\check{s}}} \left(X_{jk} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} X_{j'k} Z_{ir} \right) - E(u_{is}) \left(\sum_{\forall j' \in D_i} \frac{\partial P_{ij'|s}}{\partial \gamma_{\check{r}\check{h}\check{s}}} X_{j'k} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \beta_{rhs} \partial \beta_{\check{r}\check{k}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \beta_{\check{r}\check{k}\check{s}}} \left(X_{jk} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} X_{j'k} Z_{ir} \right) - E(u_{is}) \left(\sum_{\forall j' \in D_i} \frac{\partial P_{ij'|s}}{\partial \beta_{\check{r}\check{k}\check{s}}} X_{j'k} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \beta_{rhs} \partial \theta_{\check{t}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} \frac{\partial E(u_{is})}{\partial \theta_{\check{t}\check{s}}} \left(X_{jk} Z_{ir} - \sum_{\forall j' \in D_i} P_{ij'|s} X_{j'k} Z_{ir} \right) \\
\frac{\partial^2 \ell^2}{\partial \theta_{ts} \partial \gamma_{\check{r}\check{h}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} M_{ts} \frac{\partial E(u_{is})}{\partial \gamma_{\check{r}\check{h}\check{s}}} \\
\frac{\partial^2 \ell^2}{\partial \theta_{ts} \partial \beta_{\check{r}\check{k}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} M_{ts} \frac{\partial E(u_{is})}{\partial \beta_{\check{r}\check{k}\check{s}}} \\
\frac{\partial^2 \ell^2}{\partial \theta_{ts} \partial \theta_{\check{t}\check{s}}} &= \sum_{j=1}^J \sum_{i \in A_j} M_{ts} \left(\frac{\partial E(u_{is})}{\partial \theta_{\check{t}\check{s}}} - \frac{\partial P_{is}}{\partial \theta_{\check{t}\check{s}}} \right)
\end{aligned}$$

Y las derivadas utilizadas en las condiciones de segundo orden:

$$\begin{aligned}
\frac{\partial P_{ij'|s}}{\partial \gamma_{\check{r}\check{h}\check{s}}} &= \mathbb{1}_{(s=\check{s})} P_{ij'|s} \left(B_{ij'\check{h}} Z_{i\check{r}} - \sum_{\forall j'' \in D_i} P_{ij''|s} B_{ij''\check{h}} Z_{i\check{r}} \right) \\
\frac{\partial P_{ij'|s}}{\partial \beta_{\check{r}\check{k}\check{s}}} &= \mathbb{1}_{(s=\check{s})} P_{ij'|s} \left(X_{j'\check{k}} Z_{i\check{r}} - \sum_{\forall j'' \in D_i} P_{ij''|s} X_{ij''\check{k}} Z_{i\check{r}} \right) \\
\frac{\partial P_{is}}{\partial \theta_{\check{t}\check{s}}} &= M_{i\check{t}} P_{i\check{s}} (\mathbb{1}_{(s=\check{s})} - P_{is}) \\
\frac{\partial E(u_{is})}{\partial \gamma_{\check{r}\check{h}\check{s}}} &= \frac{P_{i\check{s}} P_{ij^*|\check{s}}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \left(B_{ij^*\check{h}} Z_{i\check{r}} - \sum_{\forall j' \in D_i} P_{ij'|\check{s}} B_{ij'\check{h}} Z_{i\check{r}} \right) \left(\mathbb{1}_{(s=\check{s})} - \frac{P_{is} P_{ij^*|s}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \right) \\
\frac{\partial E(u_{is})}{\partial \beta_{\check{r}\check{k}\check{s}}} &= \frac{P_{i\check{s}} P_{ij^*|\check{s}}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \left(X_{j^*\check{k}} Z_{i\check{r}} - \sum_{\forall j' \in D_i} P_{ij'|\check{s}} X_{j'\check{k}} Z_{i\check{r}} \right) \left(\mathbb{1}_{(s=\check{s})} - \frac{P_{is} P_{ij^*|s}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \right) \\
\frac{\partial E(u_{is})}{\partial \theta_{\check{t}\check{s}}} &= M_{i\check{t}} \frac{P_{i\check{s}} P_{ij^*|\check{s}}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \left(\mathbb{1}_{(s=\check{s})} - \frac{P_{is} P_{ij^*|s}}{\sum_{s'=1}^S P_{is'} P_{ij^*|s'}} \right)
\end{aligned}$$

G.3 Tercera Etapa

Se realizó una tercera etapa de estimación para poder incluir características dicotómicas de las alternativas altamente correlacionadas con los atributos que definen las clases, extendiendo la metodología propuesta por Bhat (1997). Fue necesario debido a que la gran flexibilidad del modelo genera que, al agregar estas variables, tiendan a ajustarse clases de tamaños suficientes para que estos atributos se vuelven clasificadores perfectos (todos los individuos dentro de la clase eligen opciones con el mismo valor el atributo). Lo anterior implica que se estimen parámetros de preferencia dentro de las clases lo suficientemente grandes para hacer del proceso de estimación inestable, pues sus valores “explotan”, imposibilitando un correcto cálculo de las probabilidad y la finalización del proceso de ajuste⁵.

⁵El caso concreto en este trabajo es la variable relacionada con enseñanza técnico profesional, elección fuertemente correlacionada al ingreso familiar y educación de los padres. A pesar de intentar diferentes combinaciones de disponibilidad de estos colegios, o discriminación de tipo de enseñanza según ingreso; inevitablemente el modelo ajusta una clase con ciertas observaciones tal que ninguna (o todas) elige colegios

La estrategia propuesta comprende utilizar los resultados de las dos primeras etapas: los parámetros de pertenencia a clase y las preferencias por los atributos de las alternativas incluidos al principio (los que no consideran estas características dicotómicas correlacionadas con atributos de pertenencia a clases). El procedimiento sigue los siguientes pasos:

1. Se realiza la primera y segunda etapa, omitiendo las características dicotómicas de las alternativas altamente correlacionadas con los atributos que definen las clases.
2. Se agregan las variables omitidas a la base utilizada en la estimación anterior.
3. Se realiza la segunda etapa (optimización directa de función de verosimilitud) nuevamente en una versión modificada; utilizando como valores iniciales los parámetros ya estimados, y cero para los parámetros asociados a las nuevas variables. La diferencia es que en esta versión no se actualizan los parámetros de pertenencia a clases θ , es decir, ya no son parte de los parámetros a ajustar. Esto permite agregar variables correlacionadas con características que definen las clases sin afectar a éstas.
4. Con el resultado anterior se calcula la matriz de covarianzas en base a la especificación directa del problema de maximización que incluye los parámetros de pertenencia a clases en el jacobiano (segunda etapa), pero sin realizar ninguna iteración del algoritmo de optimización, pues esto modificaría las clases.

No se encontró literatura de problemas similares al enfrentado, pero la experiencia empírica de este trabajo muestra que es una alternativa sencilla y cómoda para ingresar atributos dicotómicos de las alternativas correlacionadas con las variables de pertenencia a clases. No debiera generar problemas, pues se basa en un modelo (restringido) ya estimado, y se mantiene la propiedad de asignación endógena a las clases. La única crítica evidente es que esta asignación no es condicional al efecto de las variables omitidas sobre la elección, sí al resto de todos los otros atributos considerados para la primera etapa.

de ese tipo, estimando un set de parámetros de preferencia por colegios técnico profesionales muy grandes (“explotan” a tal punto que no es posible calcular la probabilidad).

G.4 Errores estándar

Los errores estándar se construyen a partir del inverso de la Matriz de Información, varianza asintótica de los estimadores de máxima verosimilitud. Como se muestra en (G.8), ésta se puede escribir en función de la esperanza del Hessiano de la función de log-verosimilitud.

$$\mathcal{I} = E(-\mathcal{H}) = -E \begin{pmatrix} \frac{\partial^2 \ell}{\partial \gamma_{rhs} \partial \gamma_{\tilde{r}\tilde{h}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \gamma_{rhs} \partial \beta_{\tilde{r}\tilde{k}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \gamma_{rhs} \partial \theta_{\tilde{i}\tilde{s}}} \\ \frac{\partial^2 \ell}{\partial \beta_{rks} \partial \gamma_{\tilde{r}\tilde{h}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \beta_{rks} \partial \beta_{\tilde{r}\tilde{k}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \beta_{rks} \partial \theta_{\tilde{i}\tilde{s}}} \\ \frac{\partial^2 \ell}{\partial \theta_{ts} \partial \gamma_{\tilde{r}\tilde{h}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \theta_{ts} \partial \beta_{\tilde{r}\tilde{k}\tilde{s}}} & \frac{\partial^2 \ell}{\partial \theta_{ts} \partial \theta_{\tilde{i}\tilde{s}}} \end{pmatrix} \quad (\text{G.8})$$

En (G.9) se exhibe el cálculo de la matriz de covarianzas, basado en que el negativo del Hessiano evaluado en los parámetros estimados es un estimador de la matriz de información y del inverso de la matriz de covarianzas.

$$\widehat{Cov}(\hat{\gamma}, \hat{\beta}, \hat{\theta}) = [-\mathcal{H}(\hat{\gamma}, \hat{\beta}, \hat{\theta})]^{-1} \quad (\text{G.9})$$

G.4.1 Otros errores estándares

Los errores estándares de estadísticos que son función de los parámetros estimados se construyen mediante el método delta. Sea η un estadístico definido a partir de una función $g(\gamma, \beta, \theta)$, y sea $\nabla g'$ el gradiente de dicha función, entonces la varianza de $\hat{\eta}$ se calcula como en (G.10).

$$\widehat{Var}(\hat{\eta}) = \nabla g' \widehat{Cov}(\hat{\gamma}, \hat{\beta}, \hat{\theta}) \nabla g \quad (\text{G.10})$$

Los errores de los efectos marginales, promedios de efectos marginales y de parámetros promedio son calculados con este método.

G.5 Efectos Marginales

Se calcula el efecto marginal de un aumento de un atributo del colegio elegido en la probabilidad de la elección del mismo. Esto se realiza a nivel de observación, para permitir agregar posteriormente considerando diferentes grupos.

G.5.1 Variables continuas

$$\begin{aligned}\frac{\partial P_{ij^*}}{\partial X_{j^*k'}} &= \frac{\partial \sum_{s=1}^S P_{is} P_{ij^*|s}}{\partial X_{j^*k'}} \\ &= \sum_{s=1}^S P_{is} \frac{\partial P_{ij^*|s}}{\partial X_{j^*k'}}\end{aligned}\tag{G.11}$$

De la ecuación (G.11) se observa que los efectos marginales para un *latent class logit* no son más que un promedio de los efectos marginales de varios logit multinomial (representando cada uno una clase), ponderados por la probabilidad de pertenencia a la clase condicional a los atributos de cada observación. Definiendo:

$$\%_{ijs} = \sum_{r=1}^R \sum_{h=1}^H \gamma_{rhs} B_{ijh} Z_{ir} + \sum_{r=1}^R \sum_{k=1}^K \beta_{rks} X_{jk} Z_{ir}$$

Se puede calcular el efecto marginal de un logit multinomial:

$$\begin{aligned}\frac{\partial P_{ij^*|s}}{\partial X_{j^*k'}} &= \frac{\partial \left(\frac{\exp(\%_{ij^*s})}{\sum_{j' \in D_i} \exp(\%_{ij's})} \right)}{\partial X_{j^*k'}} \\ &= \frac{\frac{\partial \exp(\%_{ij^*s})}{\partial X_{j^*k'}} \sum_{j' \in D_i} \exp(\%_{ij's}) - \%_{ij^*s} \frac{\partial (\sum_{j' \in D_i} \exp(\%_{ij's}))}{\partial X_{j^*k'}}}{\left(\sum_{j' \in D_i} \exp(\%_{ij's}) \right)^2} \\ &= \frac{\exp(\%_{ij^*s}) \frac{\partial \%_{ij^*s}}{\partial X_{j^*k'}} \sum_{j' \in D_i} \exp(\%_{ij's}) - \%_{ij^*s} \exp(\%_{ij^*s}) \frac{\partial \%_{ij^*s}}{\partial X_{j^*k'}}}{\left(\sum_{j' \in D_i} \exp(\%_{ij's}) \right)^2} \\ &= P_{ij^*|s} (1 - P_{ij^*|s}) \sum_{r=1}^R \beta_{rk's} Z_{ir}\end{aligned}$$

Luego, el efecto marginal que interesa es:

$$\frac{\partial P_{ij^*}}{\partial X_{j^*k'}} = \sum_{s=1}^S \left(P_{is} P_{ij^*|s} (1 - P_{ij^*|s}) \sum_{r=1}^R \beta_{rk's} Z_{ir} \right)$$

De forma análoga se puede construir $\frac{\partial P_{ij^*}}{\partial B_{ij^*h'}}$:

$$\frac{\partial P_{ij^*}}{\partial B_{ij^*h'}} = \sum_{s=1}^S \left(P_{is} P_{ij^*|s} (1 - P_{ij^*|s}) \sum_{r=1}^R \gamma_{rh's} Z_{ir} \right)$$

G.5.2 Variables binarias

En el caso de variables binarias, el efecto marginal se calcula como la diferencia en probabilidad cuando el atributo toma el valor uno y cero:

$$\frac{\partial P_{ij^*}}{\partial X_{j^*k'}} = P_{ij^*|X_{j^*k'}=1} - P_{ij^*|X_{j^*k'}=0}$$

$$\frac{\partial P_{ij^*}}{\partial B_{ij^*h'}} = P_{ij^*|B_{ij^*h'}=1} - P_{ij^*|B_{ij^*h'}=0}$$

G.6 Ajuste

La estimación con las variables definitivas se realizó para especificaciones de 3 a 8 clases. Para decisión del numero de clases se utilizó el Criterio de Información Bayesiana (BIC)⁶, aunque en la literatura relacionada a modelos de clase latente el Criterio de Información Akaike (AIC)⁷ es también popular. Una de las diferencias entre ambos es que el primero penaliza más el numero de parámetros considerados. Se exhiben en la Tabla G.1 los indicadores usuales de ajuste para modelos de elección discreta estimados por máxima verosimilitud (Ortúzar y Willumsen, 2011).

Se calculó la razón de verosimilitud considerando el modelo de igual probabilidad como modelo restringido. También los criterios de ajuste ρ^2 , $\bar{\rho}^2$ y el ρ_{adj}^2 . Por último, se calcularon estadísticos como el FPR (First Preference Recovery), CR (Chance Recovery),

⁶ $BIC = \log(n)k - 2\ell$, n : numero de observaciones

⁷ $AIC = 2k - 2\ell$, k : numero de parámetros estimados.

MSR (Market Share Recovery) y ER (Expected FPR). Ortúzar y Willumsen (2011) indican que si FPR, CR y ER son parecidos, entonces el modelo es razonable pero poco informativo, si FPR y ER son similares y mayores que CR entonces también es informativo y si FPR y ER son diferentes entonces el modelo no explica la variabilidad de los datos.

Siguiendo a Ortuzar 2011, el modelo es informativo, y pasa los test de razón de verosimilitud cuando se compara con el modelo de igual probabilidad. El modelo presenta un pseudo r-squared de 0.316

TABLA G.1. Estadísticos de ajuste del modelo

Estadístico	
$\ell(\hat{\alpha})$	-1.32e+05
$\ell(\hat{\alpha})$ promedio	-3.47
$\ell(\hat{\alpha})$ promedio (fullDisp)	-4.18
$-2[\ell(0) - \ell(\hat{\alpha})]$	1.22e+05
ρ^2	0.318
ρ_{adj}^2	0.316
FPR	0.202
MSR	0.00252
BIC	2.67e+05
AIC	2.64e+05
CR	0.00711
ER	0.21
var(CR)	0.00704
var(ER)	0.141

Una estrategia alternativa para evaluar ajuste fue probar su capacidad de predicción sobre la muestra de estimación. Una aproximación fue evaluar la dependencia del colegio predicho. En la Tabla G.2 se muestra el cruce, las columnas representan colegios cuya elección real es municipal, subvencionado y privado respectivamente, mientras que las filas las dependencias predichas para esos colegios. Se aprecia que casi un 80 % de las observaciones que en la realidad eligen privado, su elección predicha también es un colegio privado. Estos porcentajes bajan 73 % para el caso de los subvencionados y bordean

el 60 % para los municipales. Es interesante que solo un 1.3 % de las elecciones reales de colegio municipal fue predicha dentro de las alternativas privadas (y un 40 % dentro de las subvencionadas). En cambio casi un 10 % de elecciones reales en colegios privados fueron predichas dentro del mundo municipal.

Otro método consistió en verificar la cantidad de predicciones correctas según dependencia. La Tabla G.3 muestra las predicciones correctas según las primeras tres preferencias por alternativa que entrega el modelo para cada observación. A pesar de la gran cantidad de posibilidades de colegios, sobre un 19 % de las predicciones calzan con su colegio real en primera preferencia, mientras que casi un 17 % entre la segunda y tercera.

TABLA G.2. Dependencia de la predicción vs dependencia de elección real

Predicción	Elección Real			Total
	Municipal	Subvencionado	Particular	
Municipal	6490 (61.4 %)	4695 (22.3 %)	498 (8.0 %)	11683 (30.9 %)
Subvencionado	3940 (37.3 %)	15504 (73.6 %)	719 (11.6 %)	20163 (53.3 %)
Particular	138 (1.3 %)	870 (4.1 %)	5005 (80.4 %)	6013 (15.9 %)
Total	10568 (100.0 %)	21069 (100.0 %)	6222 (100.0 %)	37859 (100.0 %)

TABLA G.3. Correcta predicción entre tres primeras preferencias

	Municipal	Subvencionado	Particular	Total
1ra Preferencia	2393 (22.6 %)	4218 (20.0 %)	1031 (16.6 %)	7642 (20.2 %)
2da Preferencia	1216 (11.5 %)	2225 (10.6 %)	568 (9.1 %)	4009 (10.6 %)
3ra Preferencia	771 (7.3 %)	1460 (6.9 %)	370 (5.9 %)	2601 (6.9 %)

H. GRÁFICOS DE EFECTOS MARGINALES

Se construyeron dos tipos de gráficos a partir de los ingresos marginales, ambos evidencia de la heterogeneidad en preferencia estimada. La Figura H.1 corresponde a histogramas de los efectos marginales contruidos a nivel de observación, en que el color azul representa significancia estadística (5 %) de ser diferentes de cero, rojo lo contrario. La Figura H.2 gráfica mediante *box-plots* las distribuciones de los efectos marginales agrupados por observaciones pertenecientes al mismo decil de ingreso per cápita. La caja corresponde al rango intercuartil, las líneas que se extienden fuera de las cajas representan el primer y último decil de la distribución, el asterisco la media, y la línea dentro de la caja la mediana.

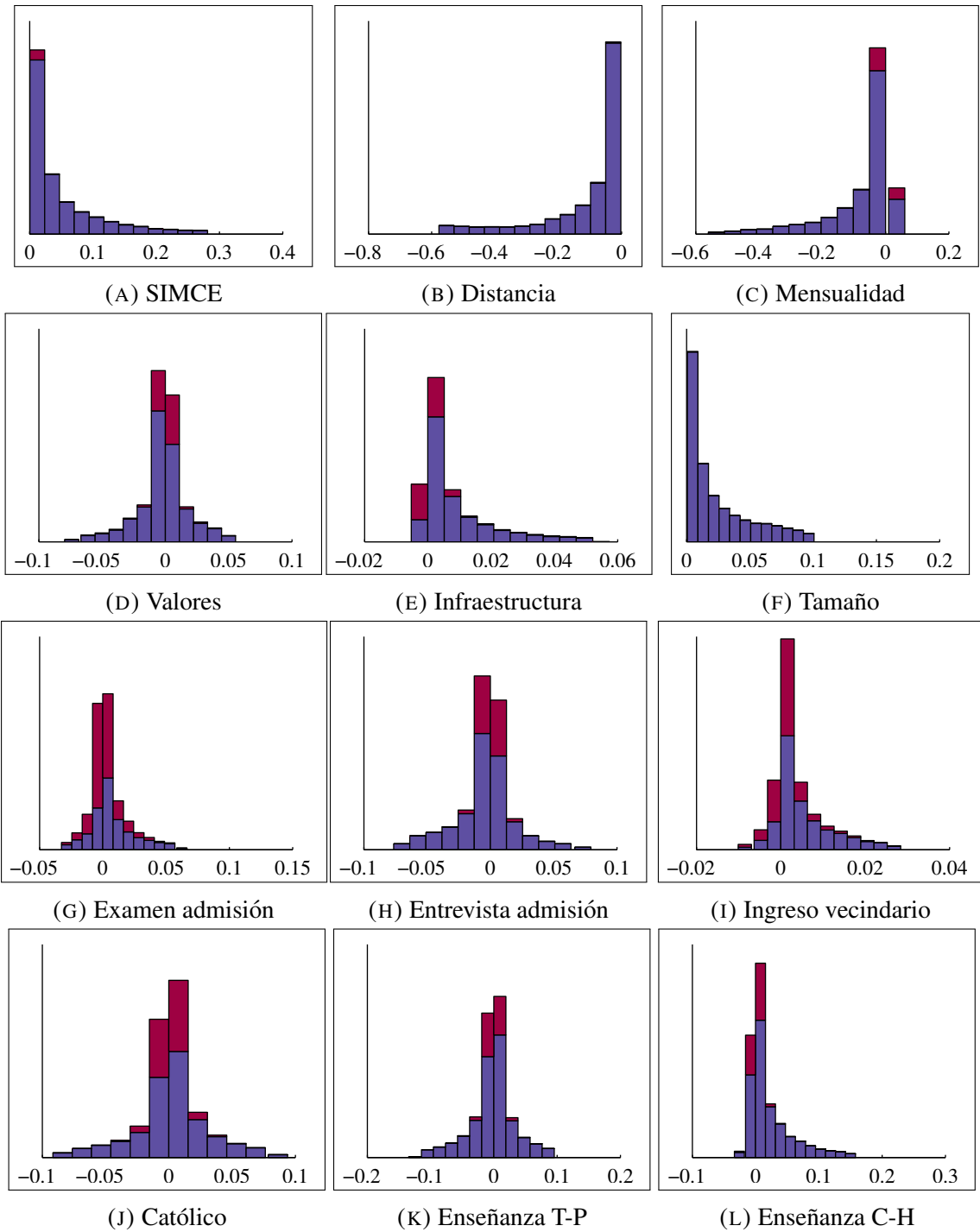


FIGURA H.1. Histogramas de efectos marginales a nivel de observación

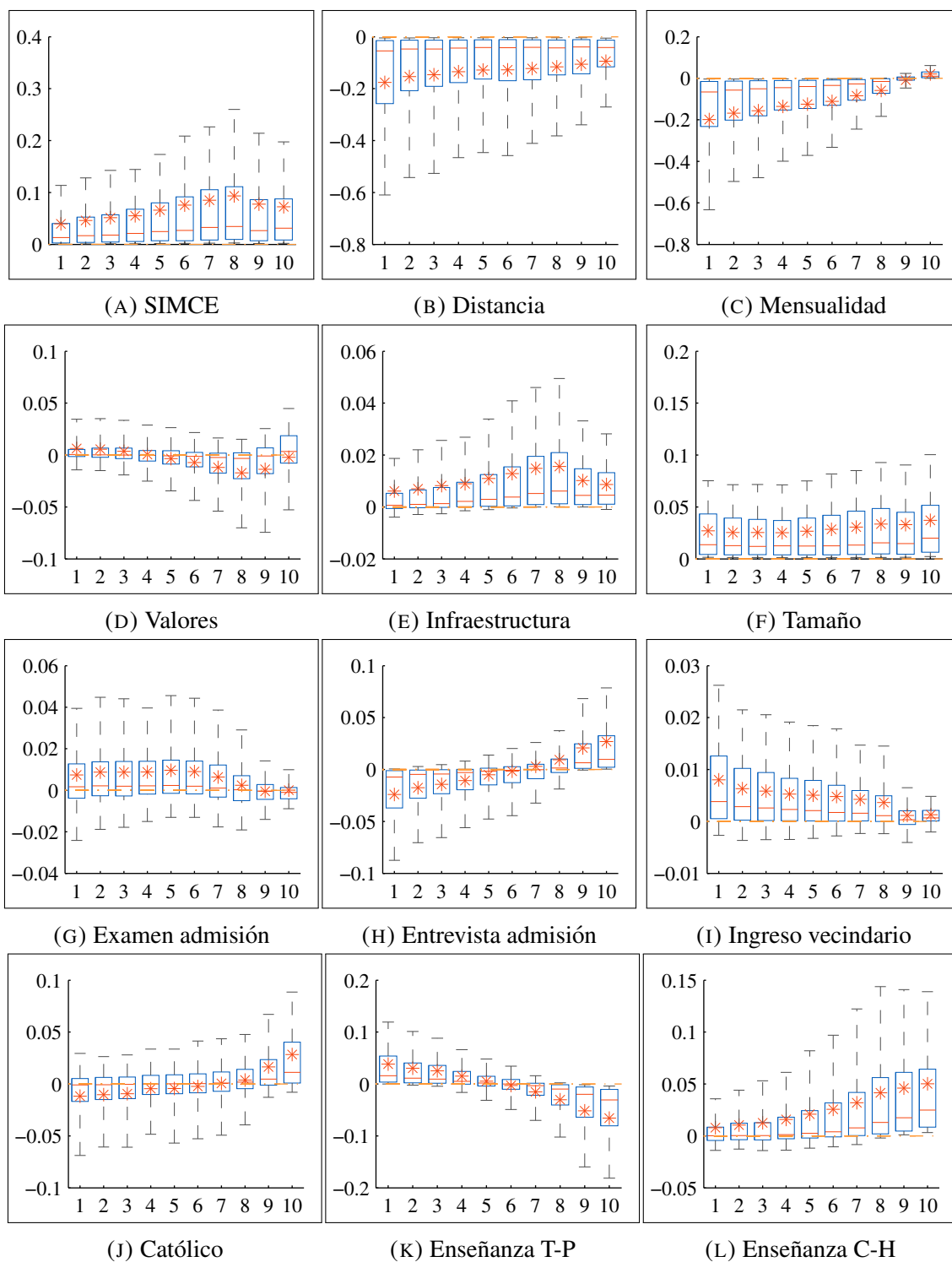


FIGURA H.2. Efectos marginales según decil de Ingreso

I. EJERCICIOS

1. Colegio original.

Observaciones asignadas en su colegio original.

2. Colegio predicho.

Observaciones asignadas en el colegio predicho. Con este escenario se comparan los otros para evaluar cambios en segregación. También cumple la función de evaluar la calidad de la predicción, pues un buen modelo debiera entregar valores de segregación cercanos a los del escenario uno (real).

3. Colegio predicho según distancia.

Observaciones distribuidas en el colegio más cercano¹. Se utiliza como *proxy* de la segregación residencial. Para este ejercicio no se respetó la cantidad de cupos por colegio.

4. Colegio predicho con distancia residencia-colegio limitada.

Observaciones distribuidas en el colegio predicho, con disponibilidad de establecimientos limitada a 2km². La idea es evaluar la segregación en un contexto de distancia limitada, que permita elección pero esté fuertemente ligada a la segregación residencial.

5. Colegio predicho con reubicación de residencias y colegios.

Observaciones distribuidas en el colegio predicho, luego de reubicar las residencias y colegios de forma aleatoria³. Contrafactual que refleja la ausencia de segregación urbana.

¹ Similar a Allen (2007).

² Para los casos en que no se tenía colegios disponibles en un radio de 2km, se amplía el radio 1km, y así sucesivamente hasta tener alguna alternativa.

³ En la práctica se permutan las ubicaciones de las observaciones de forma aleatoria, es decir no se “inventan” nuevos hogares, sino que se asigna la ubicación de unas observaciones a otras. Mismo procedimiento se realiza con los colegios.

6. Colegio aleatorio.

Observaciones distribuidas en un colegio de forma aleatoria. Escenario de integración (o “práctico sin segregación”), pues considera la configuración del sistema en cuanto a tamaño de unidades, y se construye ausente de criterios que aumenten probabilidad de asignación de ciertas observaciones condicional es sus atributos⁴.

7. Colegio aleatorio, posterior a que se llenen cupos de colegios privados.

Solo las observaciones predichas según ejercicio 2 en colegios privados se mantienen, mientras que las restantes se distribuyen en colegios municipales o subvencionados de forma aleatoria. La idea es calcular los grados de segregación imponiendo integración solo en colegios con aporte estatal.

8. Colegio predicho sin copago en colegios municipales y subvencionados.

Observaciones distribuidas en el colegio predicho, luego de fijar en cero la mensualidad para colegios municipales y subvencionados⁵. Debiera orientar la magnitud del impacto en segregación de la eliminación del financiamiento compartido.

9. Colegio predicho con reducción de oferta de colegios a $\frac{2}{3}$.

Observaciones distribuidas en el colegio predicho, luego de reducir la cantidad de colegios en un un tercio, de forma aleatoria. Los cupos se ajustaron proporcionalmente para poder

⁴Carrington y Troske (1997) critican el uso del valor cero de los índices de segregación como base desde la cual se mide la integración, en especial cuando las unidades de análisis son pequeñas. La segregación, por lo general, se entiende como el ordenamiento sistemático de los individuos, por lo que proponen que tiene más sentido compararla con la esperanza del indicador en condiciones de no asignación no sistemática (o no discriminación), que se deriva de un proceso de distribución aleatoria de las observaciones en las unidades, no con el valor cero. Los autores demuestran que la segregación medida a partir de la asignación aleatoria está correlacionado negativamente con el tamaño de las unidades y con la proporción que represente el grupo minoritario, poco tiene que ver con la cantidad de unidades.

⁵También se fijó en cero las variables relacionadas con requisitos del proceso de admisión relacionados con selección no académica (selecEntrevista y selecReligioso) para los colegios en que se eliminó el copago.

atender la demanda. Sirve para evaluar el impacto en medición de segregación del tamaño y cantidad de unidades.

10. Colegio predicho sin selección académica en colegios municipales y subvencionados.

Observaciones distribuidas en el colegio predicho, luego de quitar la restricción en disponibilidad según criterio de selección académica para colegios municipales y subvencionados⁶. Refleja una política de eliminar la selección por notas, política presente en gran parte de los liceos de mejor rendimiento.

11. Colegio predicho, observaciones con igual valoración del tipo de enseñanza.

Observaciones distribuidas en el colegio predicho, luego de imponer que no existe preferencia por algún tipo de currículum de enseñanza (técnico profesional o científico-humanista).

12. Colegio predicho, observaciones con igual valoración del SIMCE.

Observaciones distribuidas en el colegio predicho, luego de imponer que la preferencia por SIMCE sea igual para todos⁷.

13. Colegio predicho, observaciones con igual costo de la distancia.

Observaciones distribuidas en el colegio predicho según el modelo de elección, luego de imponer que el costo marginal atribuido a la distancia sea igual para todos.

14. Colegio predicho, observaciones con igual valoración de otros atributos.

⁶Se fijó en cero también las variables relacionadas con requisitos del proceso de admisión relacionados con selección académica (selecExamen) para los colegios que se les eliminó la restricción de disponibilidad.

⁷El modelo estimado permite que cada individuo tenga preferencia distinta por los atributos de los colegios, de acuerdo a sus propias características. En este ejercicio se impone que todos tengan la preferencia promedio por puntaje SIMCE.

Observaciones distribuidas en el colegio predicho según el modelo de elección, luego de imponer que el costo marginal atribuido a la todos los atributos distintos al copago, distancia, SIMCE y currículum sea igual para todos.

15. **Colegio predicho sin copago en colegios municipales y subvencionados, con reubicación de residencias y colegios.**

Combinación de los ejercicios 5 y 8.

16. **Colegio predicho, observaciones con igual valoración de tipo de enseñanza, SIMCE y otros atributos**

Combinación de ejercicios 11, 12 y 14.

17. **Colegio predicho, observaciones con igual valoración de tipo de enseñanza, SIMCE y otros atributos, con reubicación de residencias y colegios, y sin copago en colegios municipales y subvencionados.**

Combinación de ejercicios 5, 8, 11, 12 y 14.

18. **Colegio predicho, observaciones con igual valoración de tipo de enseñanza, SIMCE y otros atributos, con reubicación de residencias y colegios, sin copago en colegios municipales y subvencionados, e igual costo de la distancia.**

Combinación de ejercicios 5, 8, 11, 12, 13 y 14.

19. **Colegio predicho, observaciones con igual valoración de tipo de enseñanza, SIMCE y otros atributos, con reubicación de residencias y colegios, sin copago en colegios municipales y subvencionados, y sin selección académica en colegios municipales y subvencionados.**

Combinación de ejercicios 5, 8, 10, 11, 12 y 13.

20. **Colegio predicho, observaciones con igual valoración de tipo de enseñanza, SIMCE y otros atributos, con reubicación de residencias y colegios, sin copago en colegios municipales y subvencionados, sin selección académica en colegios municipales y subvencionados, e igual costo de la distancia.**

Combinación de ejercicios 5, 8, 10, 11, 12, 13 y 14.

J. REGLAS DE ASIGNACIÓN DE COLEGIO SEGÚN PREDICCIÓN

El modelo predice una probabilidad de elección para cada observación. Para poder definir un solo colegio como el predicho se consideró los tres colegios con mayor probabilidad de elección¹, y entre éstos se eligió de manera aleatoria, con probabilidad de elección ponderada por la probabilidad predicha por el modelo.

Para las observaciones que no quedan asignadas en la primera iteración por sobrede-manda del colegio predicho, en la siguiente iteración se buscó los tres colegios con cupo disponible con mayor probabilidad predicha, y se eligió entre éstos con la misma metodología anterior. El algoritmo se detiene cuando (1) ya están todos distribuidos en algún colegio, respetando la capacidad del colegio, (2) si de una iteración a otra no disminuye la cantidad de observaciones sin asignación, o (3) si llega las 100 iteraciones². Cuando se detuvo sin asignar a todos, las observaciones sin colegios fueron distribuidas en el último colegio predicho, utilizando un sobre cupo³.

¹Utilizando más de tres el resultado en segregación tiende a subestimar el real, aunque los efectos calculados en los contrafactuales tienen magnitud similar.

²En ningún ejercicio llegó a 100 iteraciones.

³Se realizamos corridas de manera que todos quedaran asignados, obteniendo cualitativamente los mismos resultados, pero en un tiempo mayor.